The VoicePrivacy 2020 Challenge

Challenge setup and results

Odyssey 2020

Presenter: Natalia Tomashenko

Natalia Tomashenko [1]
Brij M.L. Srivastava [2]
Xin Wang [3]
Emmanuel Vincent [2]
Andreas Nautsch [4]
Junichi Yamagishi [3,5]
Nicholas Evans [4]
Jose Patino [4]
Jean-François Bonastre [1]
Paul-Gauthier Noé [1]
Massimiliano Todisco [4]
Mohamed Maouche [2]
Benjamin O'Brien [6]
Anais Chanclu [1]

[1] LIA – University of Avignon – France
[2] Inria – France
[3] NII – Tokyo – Japan
[4] Audio Security and Privacy Group, EURECOM – France
[5] University of Edinburgh – UK
[6] Aix-Marseille University – France
Laboratoire Parole et Langage (LPL)

4th November 2020

Introduction: aim

Promote the development of privacy preservation tools for speech technology

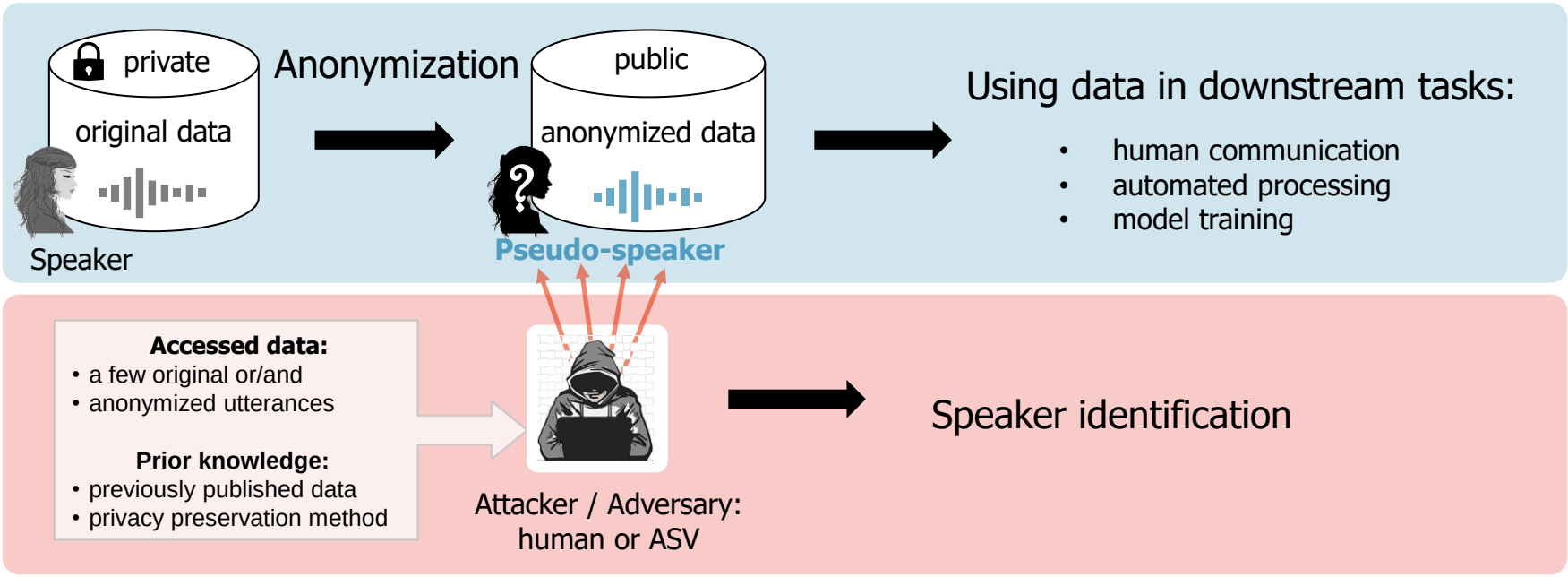
Introduction: privacy preservation for speech

- Deletion [Cohen-Hadria 2019] [Gontier 2020]
- Encryption [Pathak 2013] [Smaragdis 2007]
- Distributed learning [Leroy 2019]
- **Anonymization**
  - noise addition [Hashimoto 2016]
  - speech transformation [Qian 2017]
  - voice conversion [Jin 2009]
  - speech synthesis [Fang 2019]
  - adversarial learning [Srivastava 2019]

- ✓ suppress personally identifiable information in the speech signal
- ✓ keep unchanged all other characteristics
  - linguistic content
  - speech quality/naturalness

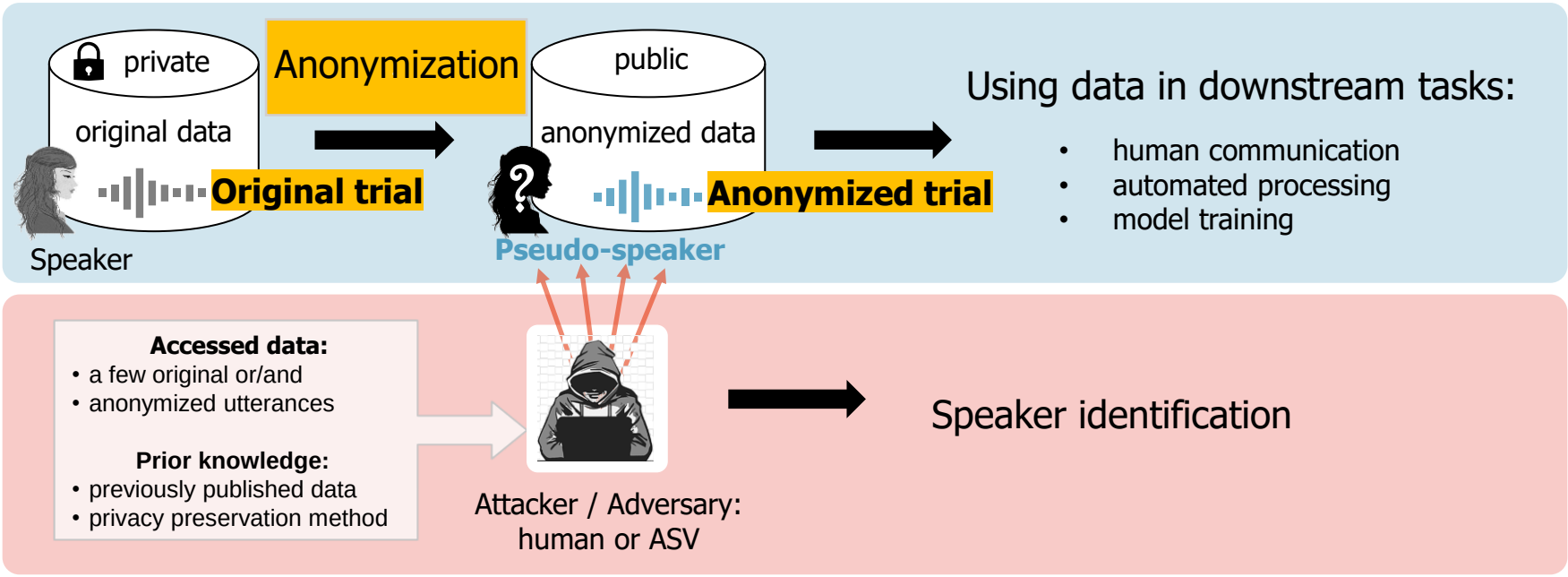
Anonymization task

- Privacy preservation is formulated as a **game** between **users** (publish some data) & **attackers** (access this data or data derived from it and wish to infer information about the users)



Anonymization task

- Privacy preservation is formulated as a **game** between **users** (publish some data) & **attackers** (access this data or data derived from it and wish to infer information about the users)



Objective evaluation: privacy and utility metrics

Privacy

ASV_{eval} Automatic speaker verification system

Objective speaker verifiability:

Equal error rate

$$\mathrm{EER} = P_{\mathrm{fa}}(\theta_{\mathrm{EER}}) = P_{\mathrm{miss}}(\theta_{\mathrm{EER}})$$

Log-likelihood-ratio cost function

$$C_{\mathrm{llr}} = \frac{\sum_{i \in \mathrm{tar}} \log_2\left(1 + e^{-\mathrm{LLR}_i}\right)}{2N_{\mathrm{tar}}} + \frac{\sum_{j \in \mathrm{imp}} \log_2\left(1 + e^{\mathrm{LLR}_j}\right)}{2N_{\mathrm{imp}}}$$

Discrimination loss

$$C_{\mathrm{llr}}^{\mathrm{min}}$$

Utility

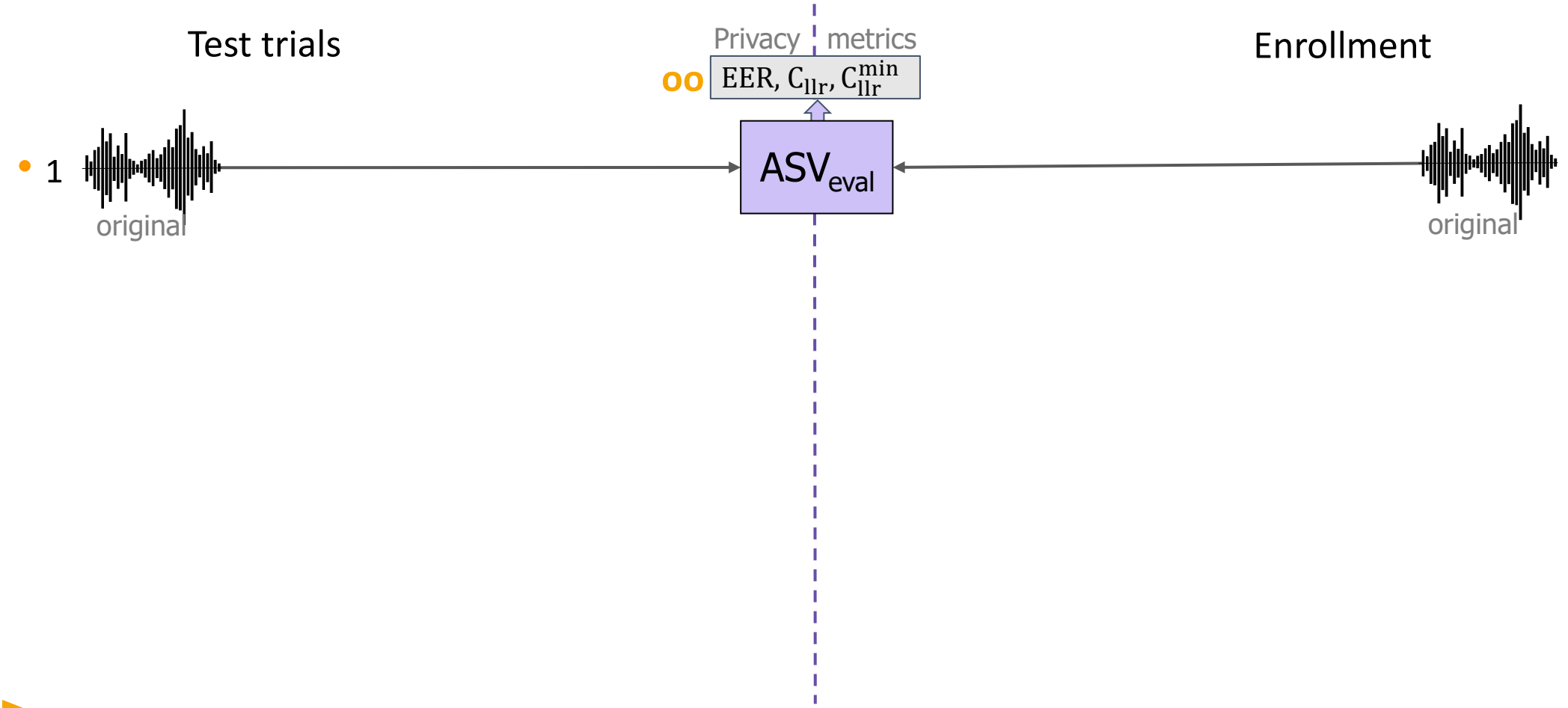
ASR_{eval} Automatic speech recognition system

Word error rate

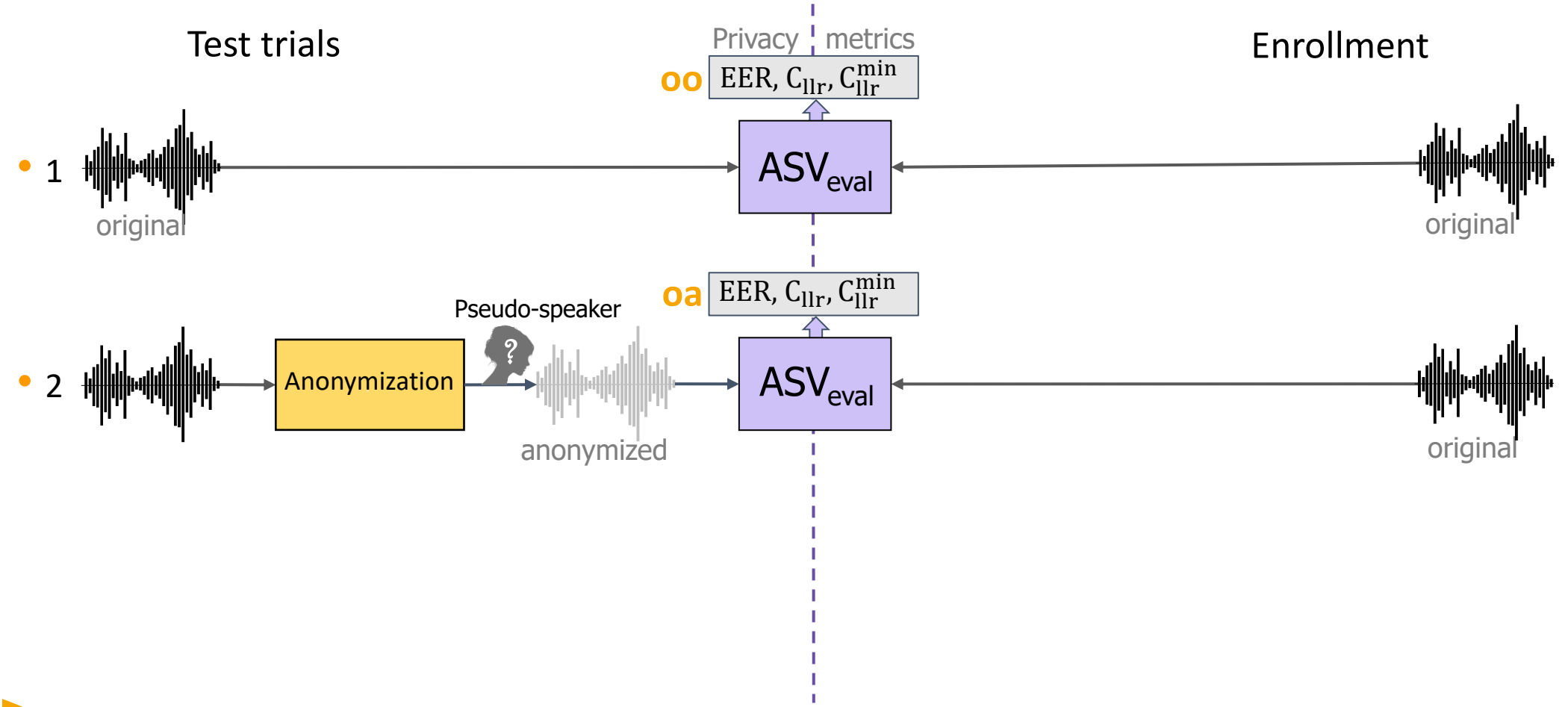
$$\mathrm{WER} = \frac{N_{\mathrm{sub}} + N_{\mathrm{del}} + N_{\mathrm{ins}}}{N_{\mathrm{ref}}}$$

Training set for ASV_{eval} & ASR_{eval}: LibriSpeech-360-clean, original data

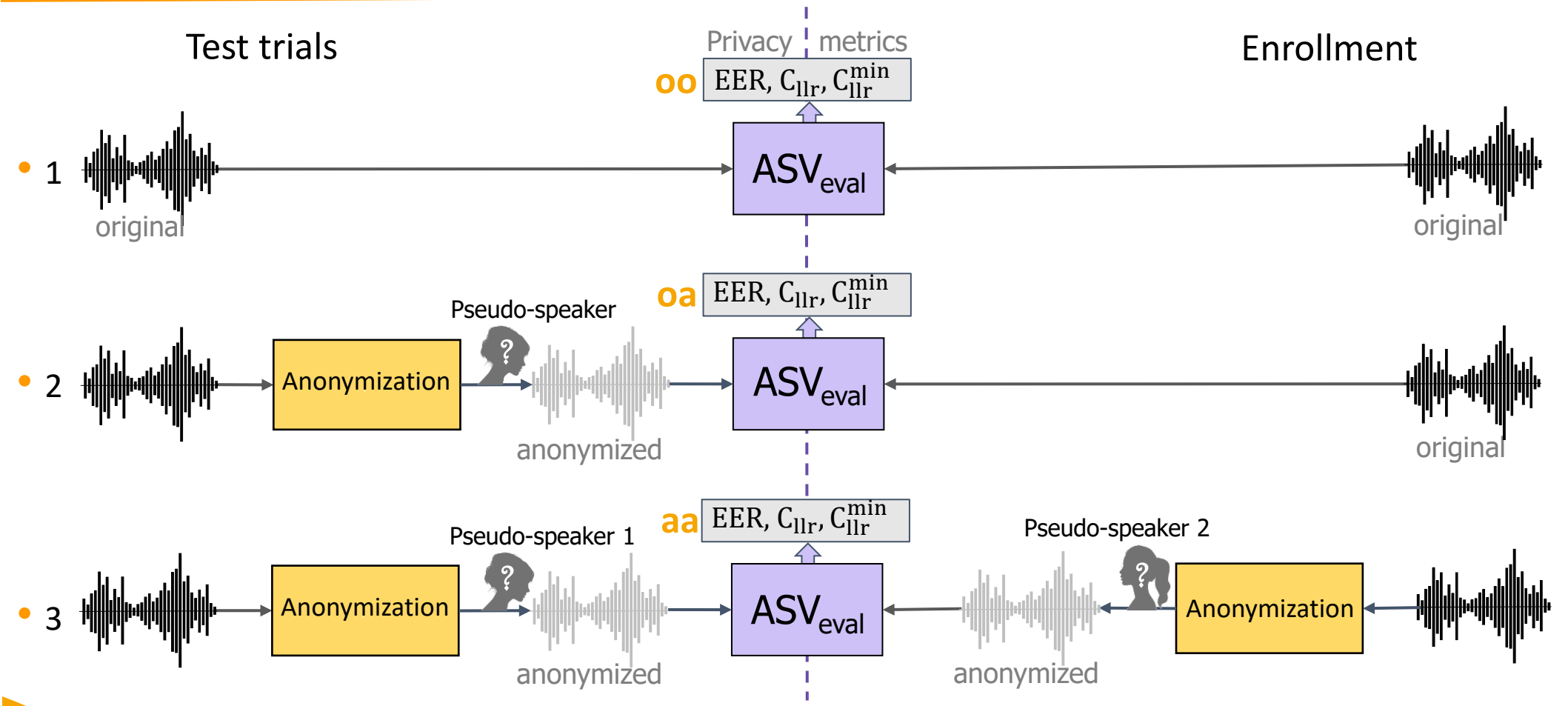
Objective evaluation: automatic speaker verification (ASV_{eval})



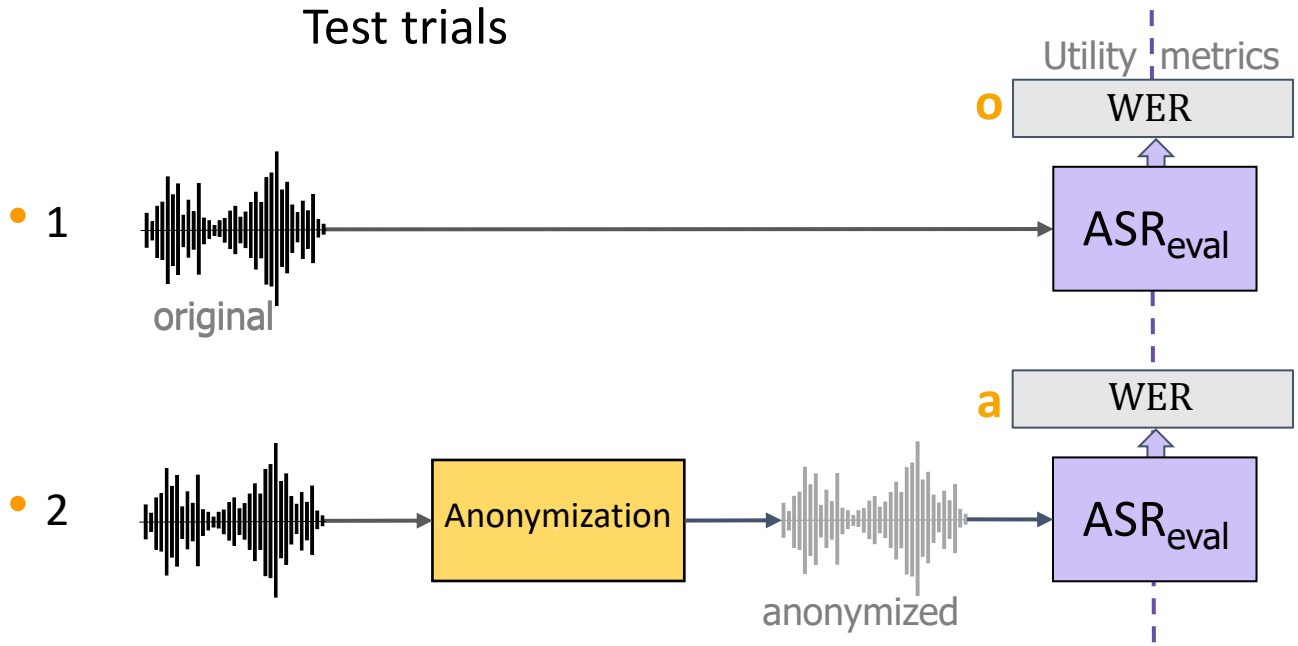
Objective evaluation: automatic speaker verification (ASV_{eval})



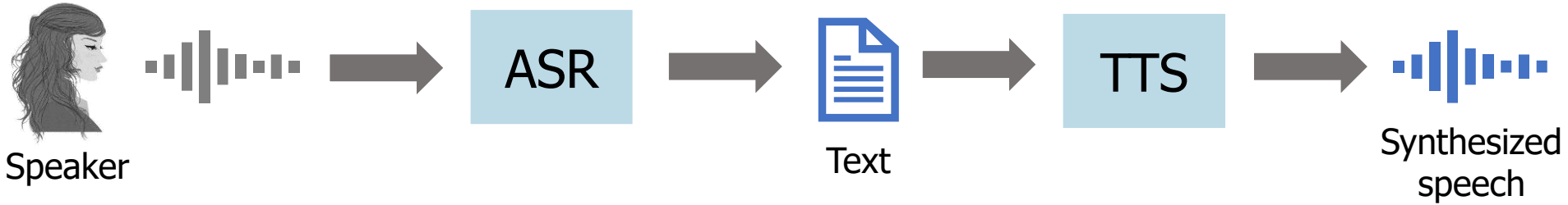
Objective evaluation: automatic speaker verification (ASV_{eval})



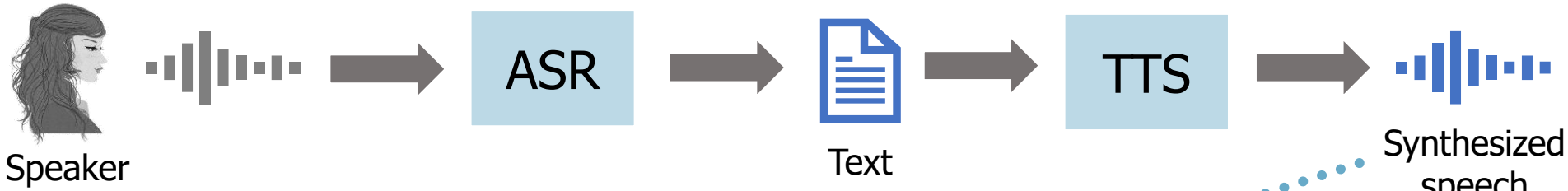
Objective evaluation: automatic speech recognition (ASR_{eval})



ASR+TTS – is it a good solution?

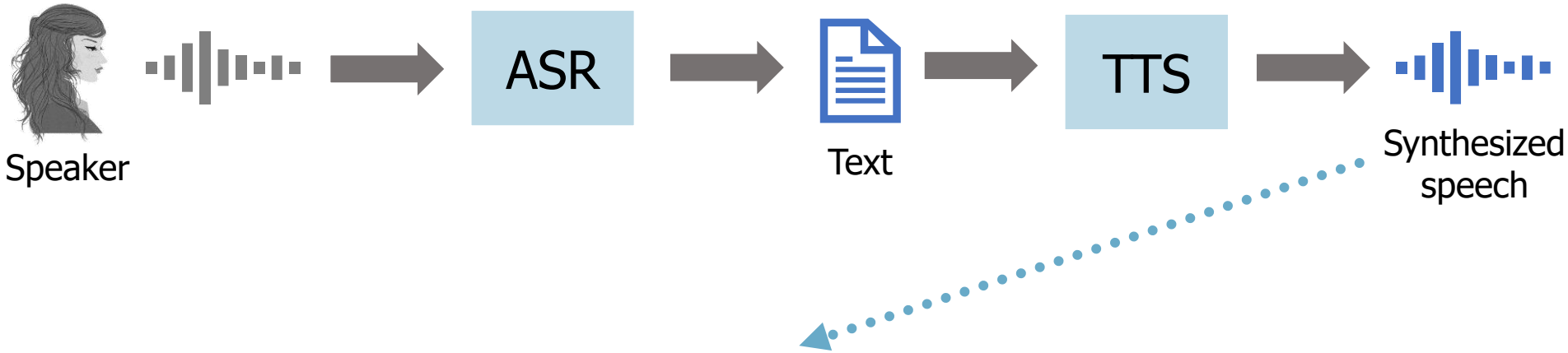


ASR+TTS – is it a good solution?



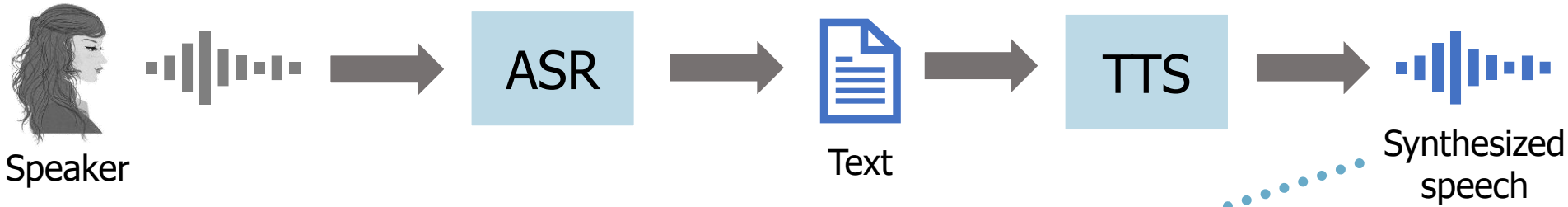
- Remove speaker identity? **yes**
- Keep unchanged all other characteristics (i. e. prosody, emotions,…)?
- Preserve linguistic content?
- Diversity and distinguishability of synthesized voices?

ASR+TTS – is it a good solution?



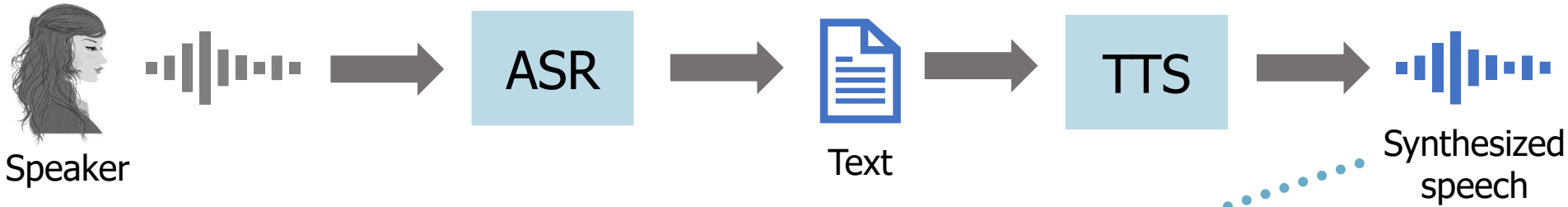
- Remove speaker identity? **yes**
- Keep unchanged all other characteristics (i. e. prosody, emotions,…)? **no**
- Preserve linguistic content?
- Diversity and distinguishability of synthesized voices?

ASR+TTS – is it a good solution?



- Remove speaker identity? **yes**
- Keep unchanged all other characteristics (i. e. prosody, emotions,…)? **no**
- Preserve linguistic content? **yes, but not perfectly due to ASR errors**
- Diversity and distinguishability of synthesized voices?

ASR+TTS – is it a good solution?



- Remove speaker identity? **yes**
- Keep unchanged all other characteristics (i. e. prosody, emotions,…)? **no**
- Preserve linguistic content? **yes, but not perfectly due to ASR errors**
- Diversity and distinguishability of synthesized voices? **limited**

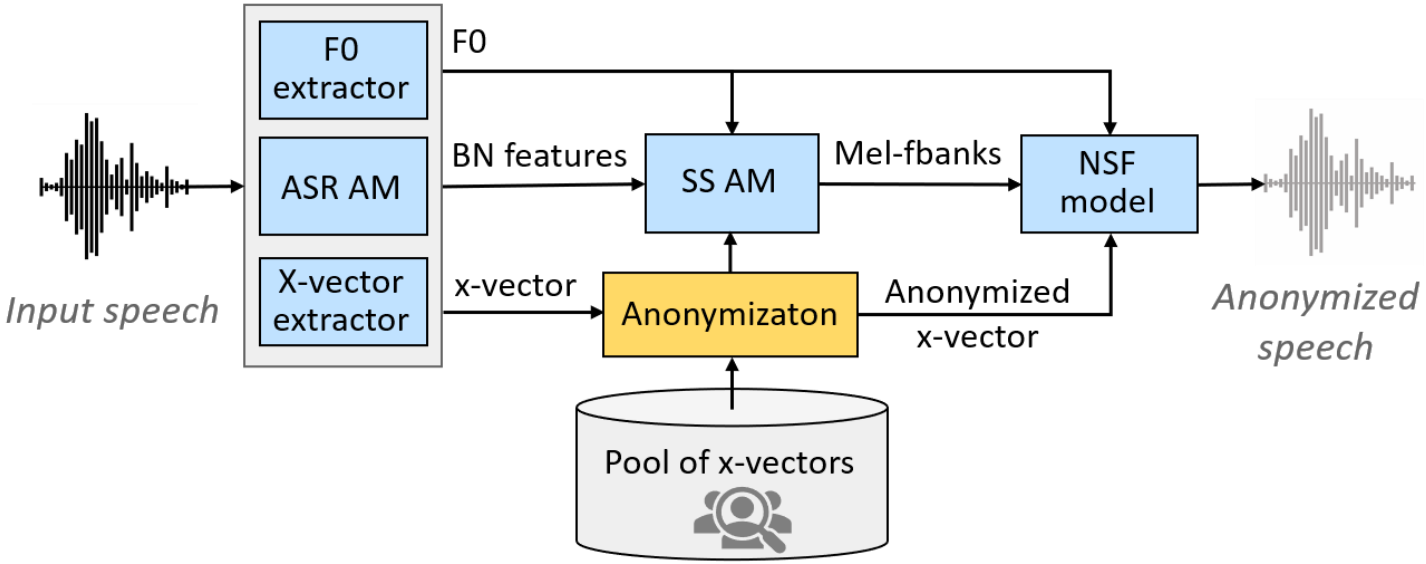
Datasets

Training		Speakers	Size, h
VoxCeleb-1,2		7363	2794
LibriSpeech:	-train-clean-100	251	100
	-train-other-500	1166	497
LibriTTS:	-train-clean-100	247	54
	-train-other-500	1160	310

Development		Speakers	Target trials	Imposter trials
LibriSpeech:	-dev-clean	29	1348	27362
VCTK-dev:	-common	30	695 3796	9721 26204
VCTK-dev:	-different			

Evaluation		Speakers	Target trials	Imposter trials
LibriSpeech:	-test-clean	29	997	20653
VCTK-test:	-common	30	700 3686	9790 26314
VCTK-test:	-different			

Baseline 1 Anonymization using x-vectors and neural waveform models

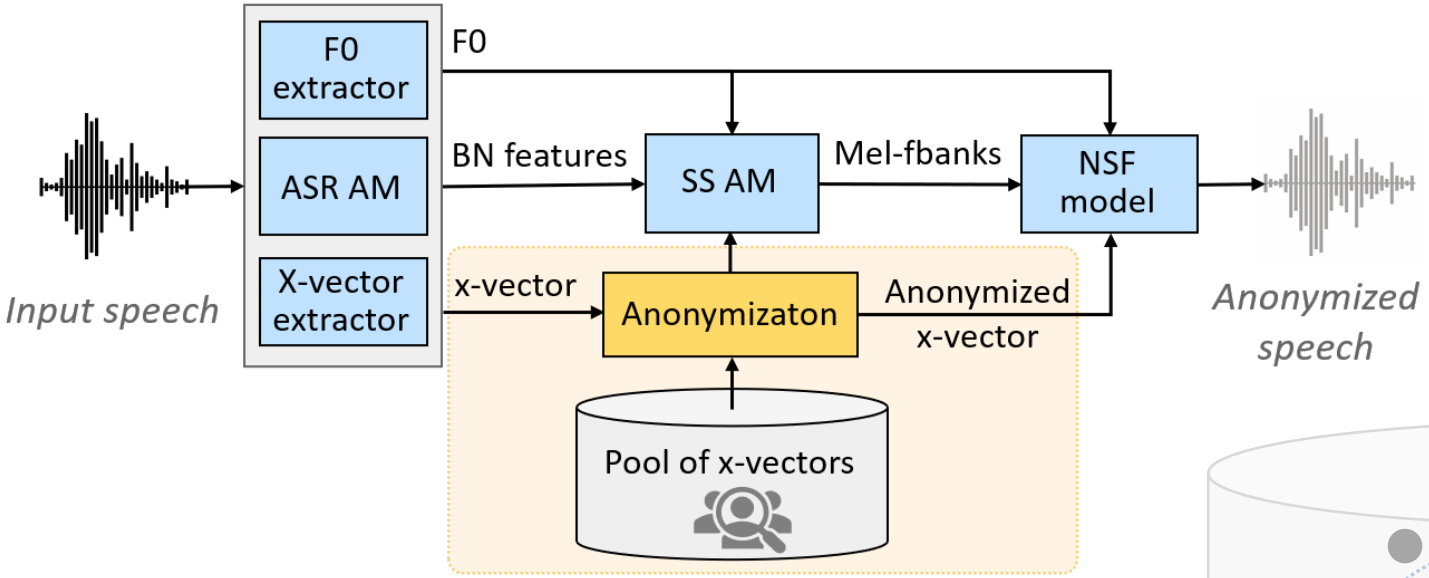


- **ASR AM**: Automatic speech recognition acoustic model (to extract **BN** (bottle-neck) features)
- **SS AM**: Speech synthesis acoustic model
- **NSF**: Neural source-filter model

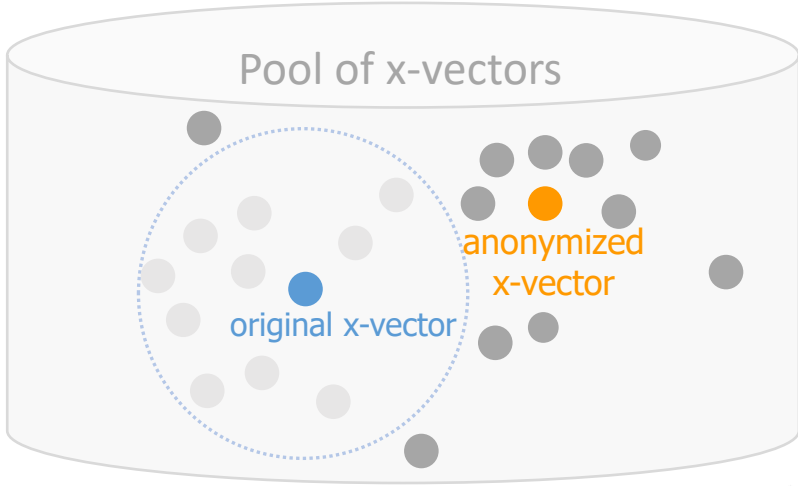
https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2020

Inspired from: [Fang 2019]

Baseline 1 Anonymization using x-vectors and neural waveform models

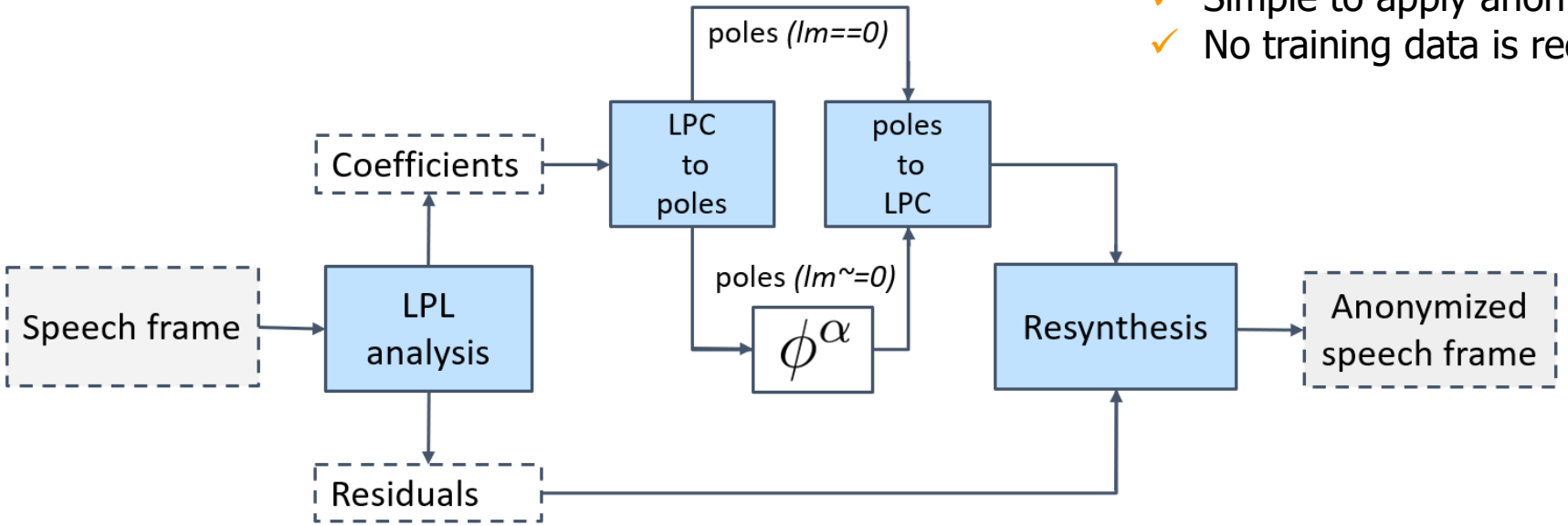


1. Choose N x-vectors farthest from the original one (**PLDA**/cosine)
2. Choose $N^* < N$ randomly from them
3. Average N^* x-vectors to obtain an anonymized x-vector



Baseline 2 Anonymization using McAdams coefficient

- ✓ Simple to apply anonymization
- ✓ No training data is required



The angle ϕ of poles with a non-zero imaginary part are raised to the power of the McAdams coefficient α to provoke a shift in frequency of its associated formant.

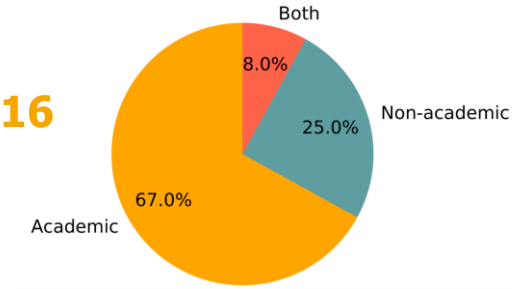
- **LPC**: linear predictive coding

https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2020

[McAdams 1984] [Patino 2020]

Participants

- Registered teams: **25** (more than **45** participants) from **13** countries
- Teams submitted valid results: **7** (**+1** contribution related to evaluation models)
  - deadline-1: submissions from **6** teams
  - deadline-2: submissions from **3** teams
- Submitted anonymization systems: **16**



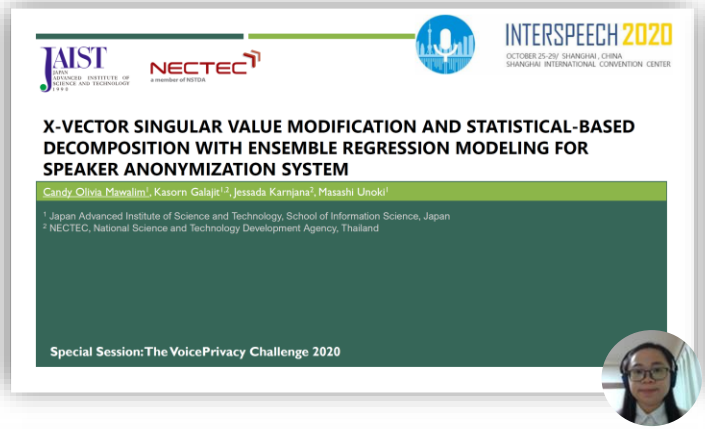
- Post-evaluation analysis (submission of the anonymized dataset for training evaluation models): **4**

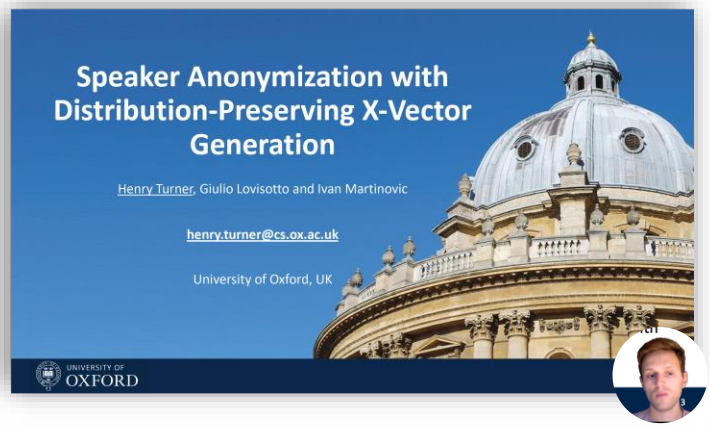
Team	Country	Status
Idiap-NKI	Switzerland	academic
Biometric Vox	Spain	non-academic
DA-IICT Speech Group	India	academic
Team SDU	Turkey	academic
PingAn	USA	non-academic
AIS-lab JAIST	Japan / Thailand	academic
BlackBox@CMU	USA	academic
Motorola Solutions	USA	non-academic
MultiSpeech	France	academic
Orange ITAAC Team	France	non-academic
Oxford System Security Lab	UK	academic
Preech	USA	academic
Sigma Technologies S.L.U.	Spain	both
TMU	Japan	academic
Ioenix	USA	non-academic
VTouch	China	academic
VIAX	China	academic
PhoneClearly.com	USA	non-academic
Kyoto Team	Japan	academic
PSUT	Jordan	academic
TJU-VP	China	academic
Late registration		
EAM AAU ANONYMOUS	Denmark	academic
TalkMeUp	USA	both
Fearghal Sheehan	Ireland	academic

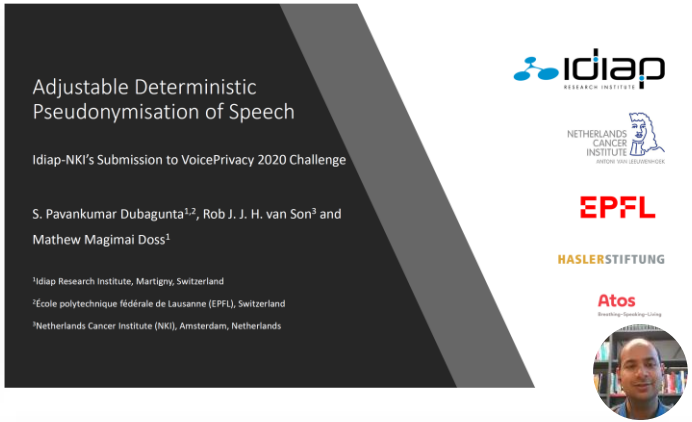
Participants' systems

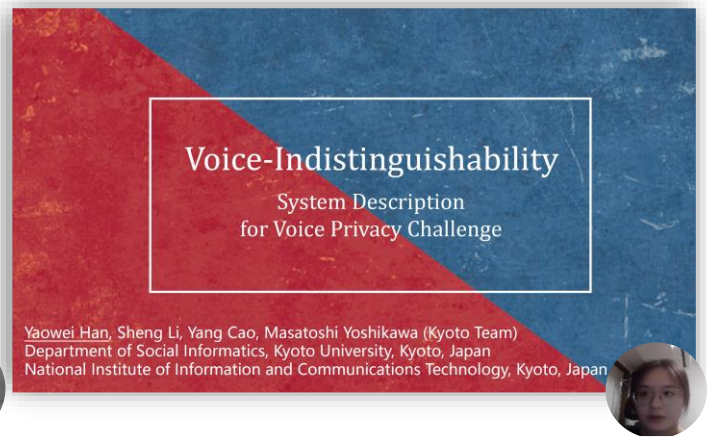
Team	Team notation	System	Deadline	System notation
AIS-lab JAIST	A	primary	1	**A1**
		primary	2	**A2**
DA-IICTSpeechGroup	D	primary	1	**D1**
Idiap-NKI	I	primary	1	**I1**
Kyoto Team	K	primary	2	**K2**
MultiSpeech	M	primary	1	**M1**
		contrastive 1	1	**M1c1**
		contrastive 2	1	**M1c2**
		contrastive 3	1	**M1c3**
		contrastive 4	1	**M1c4**
OxfordSystemSecurityLab	O	primary	1	**O1**
		contrastive 1	1	**O1c1**
Sigma_Technologies_SLU	S	primary	1	**S1**
		contrastive 1	1	**S1c1**
		primary	2	**S2**
		contrastive 1	2	**S2c1**
		baseline-1	-	**B1**
		baseline-2	-	**B2**

Participants' systems





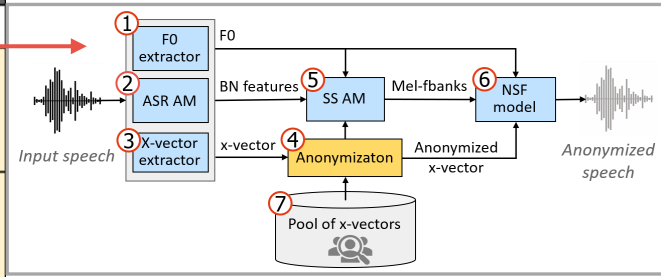




https://www.voiceprivacychallenge.org/

Participants' systems

System	Description	Modified components / data in B1							
		①	②	③	④	⑤	⑥	⑦	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500 LibriTTS train-clean-100
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling	+			+			+	
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension				+			+	Speaker pool: LibriTTS-train-other-500 VoxCeleb-1,2
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors				+			+	
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization				+				
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction				+				
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)		+			+	+		
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)		+			+	+		
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								



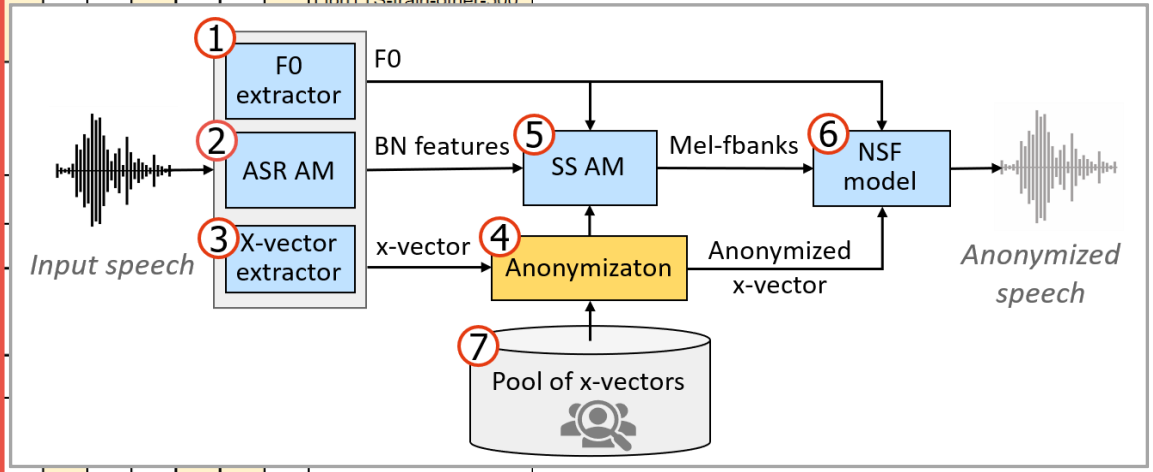
+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

Related to B1

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								



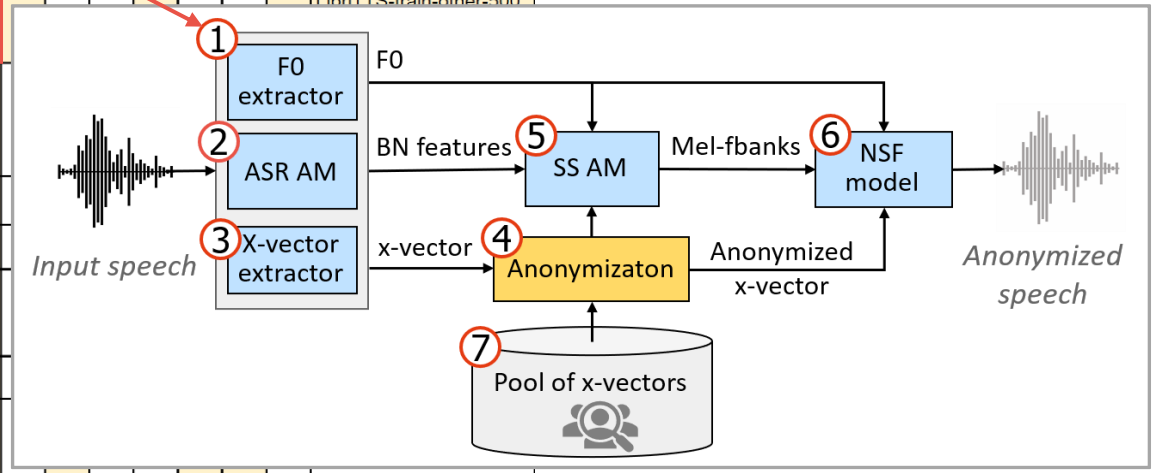
+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

Related to B1

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								



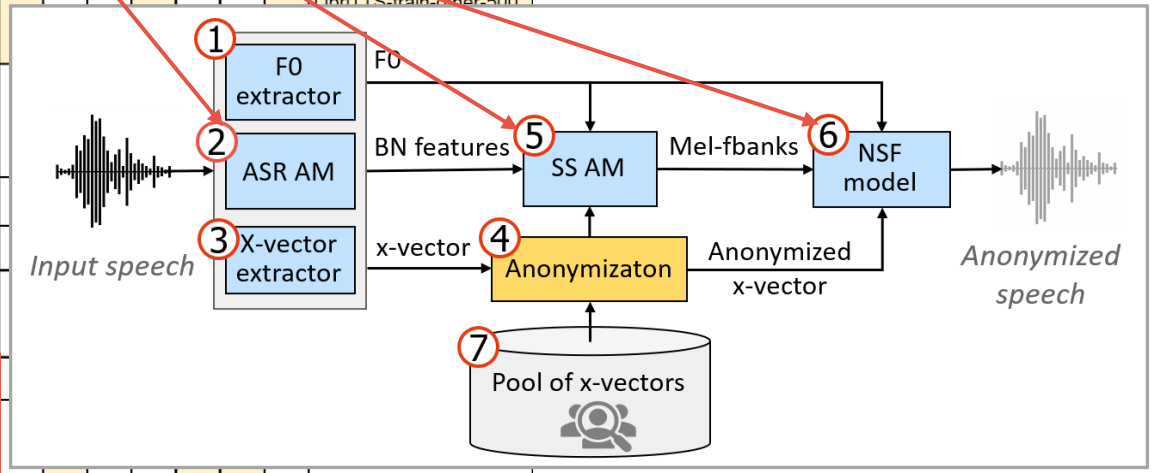
+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

Related to B1

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1:** x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1:** x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								

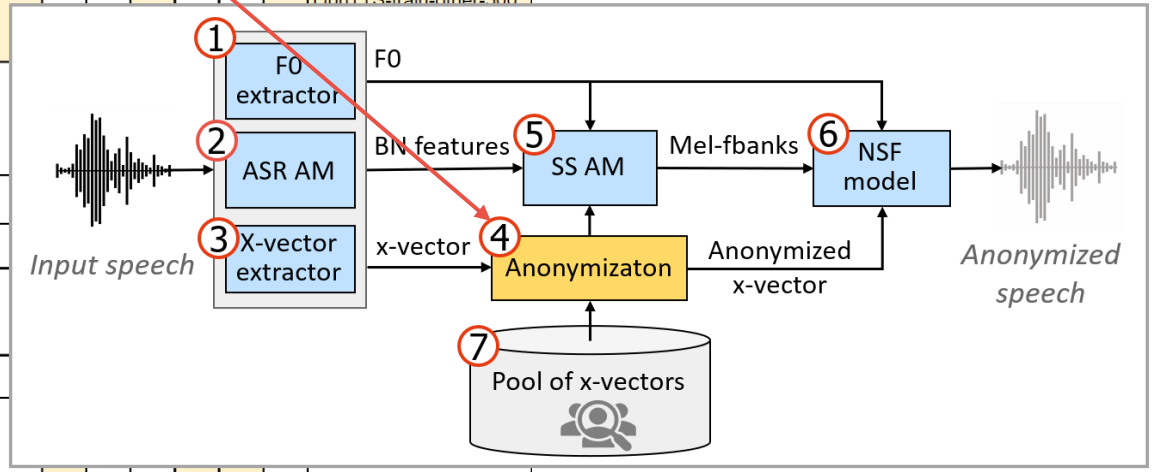


★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

Related to B1

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1:** x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1:** x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								

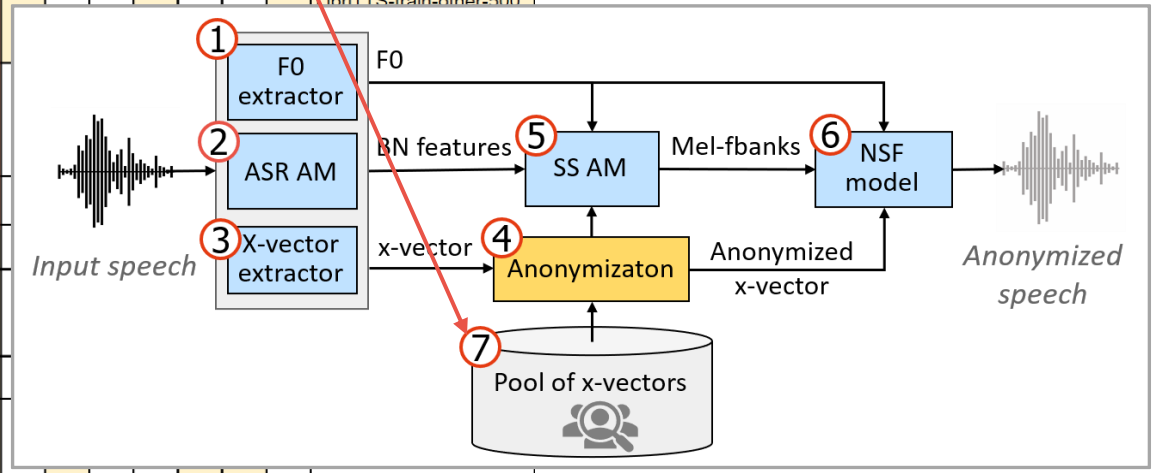


+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

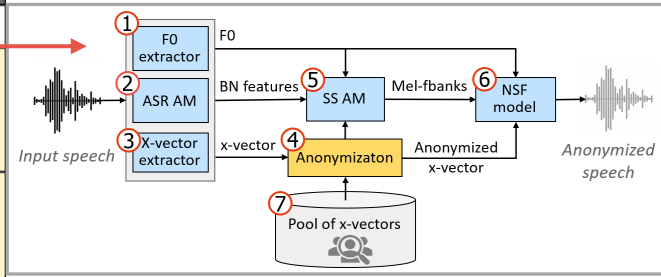
System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								



★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500 LibriTTS train-clean-100
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling	+			+			+	
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension				+			+	Speaker pool: LibriTTS-train-other-500 VoxCeleb-1,2
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors				+			+	
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization				+				
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction				+				
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)		+			+	+		
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)		+			+	+		
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								



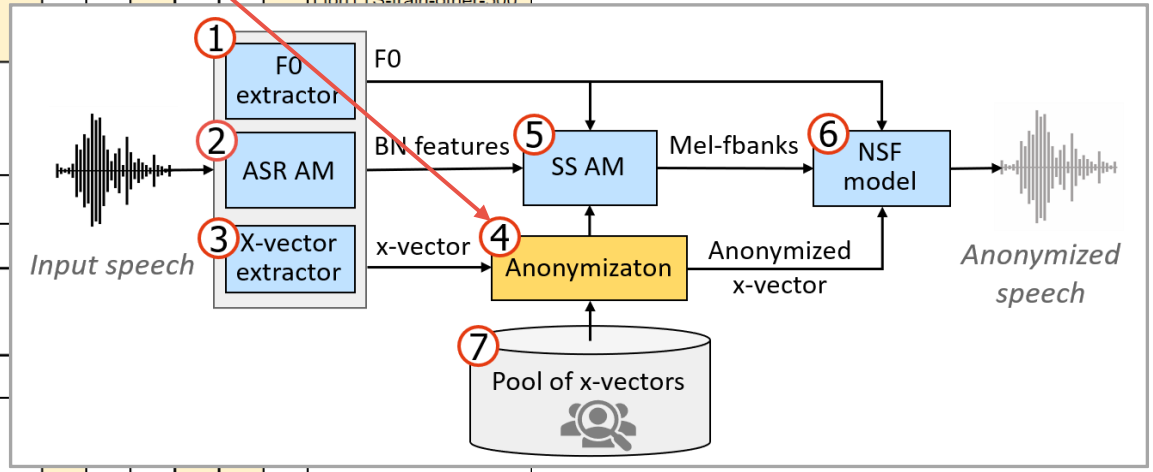
+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Participants' systems

Related to B1

System	Description	Modified components / data in B1							
		1	2	3	4	5	6	7	Data
A2	**B1**: x-vector anonymization using singular value modification				+			+	Speaker pool: LibriTTS-train-other-500
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling								
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension								
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors								
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization								
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction								
★ M1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs)								
M1c1	**B1:** ASR part to extract BN features for SS models (E2E ASR for BNs; semi-adversarial training to learn linguistic features while masking speaker information)								
M1c2	**B1:** copy-synthesis (original x-vectors)				+				
M1c3	**B1:** x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original				+				
M1c4	**B1:** x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized				+				
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm								Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius								
I1	modifications in formants, F0 and speaking rate								

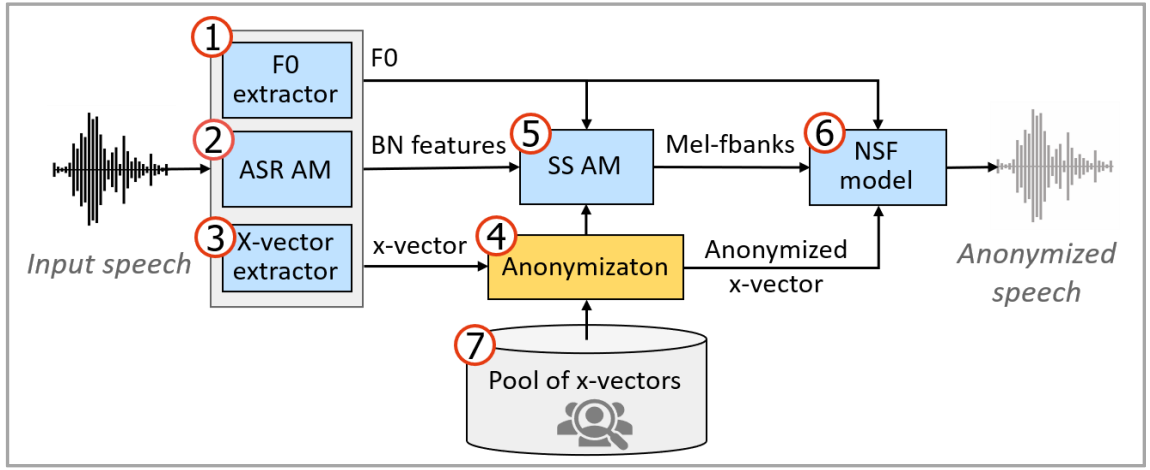


+ Modifications in B1

★ Teams submitted additional anonymized speech data for post-evaluation analysis

Approaches to x-vector anonymization

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

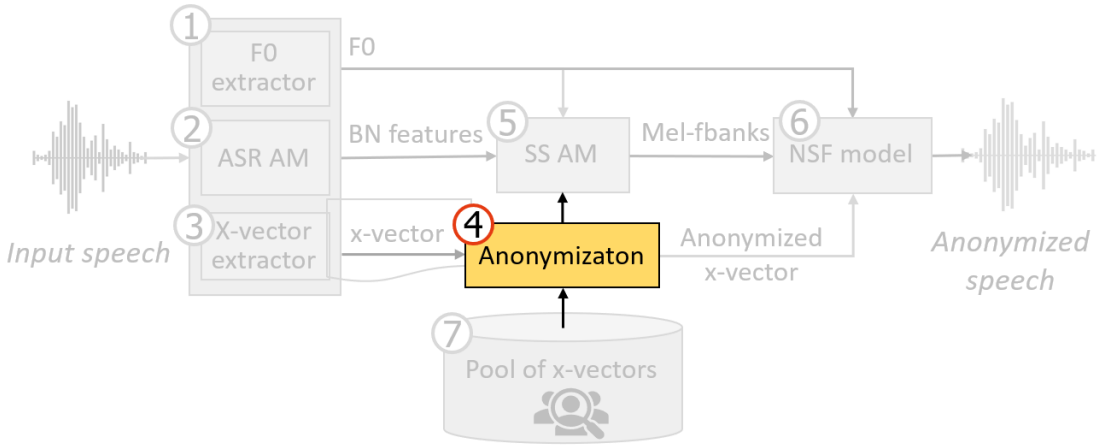


M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

Approaches to x-vector anonymization

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

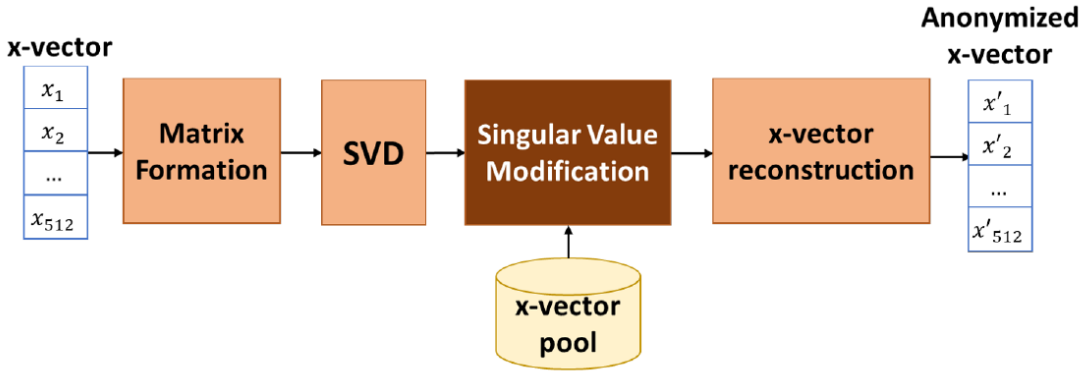


Approaches to x-vector anonymization: A2

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

A2: Singular value modification [Mawalim 2020]



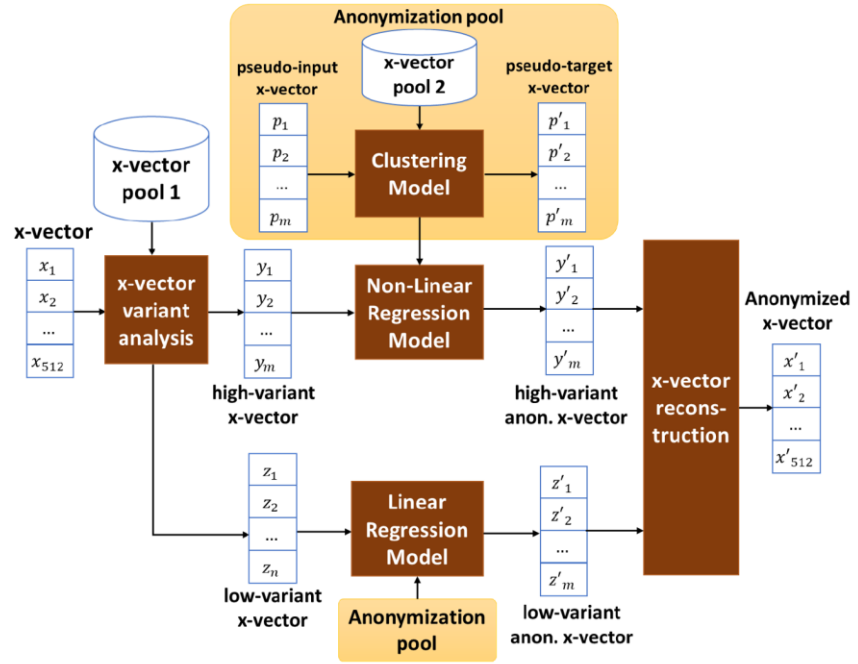
*The figure is copied from the presentation: [Mawalim 2020] X-Vector Singular Value Modification and Statistical-Based Decomposition with Ensemble Regression Modeling for Speaker Anonymization System. Candy Olivia Mawalim, Kasorn Galajit, Jessada Karnjana, Masashi Unoki

Approaches to x-vector anonymization: A

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

A: Statistical-based decomposition with regression models [Mawalim 2020]



*The figure is copied from the presentation: [Mawalim 2020] X-Vector Singular Value Modification and Statistical-Based Decomposition with Ensemble Regression Modeling for Speaker Anonymization System. Candy Olivia Mawalim, Kasorn Galajit, Jessada Karnjana, Masashi Unoki

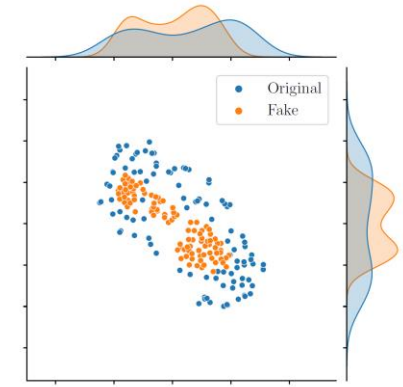
Approaches to x-vector anonymization: O1

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

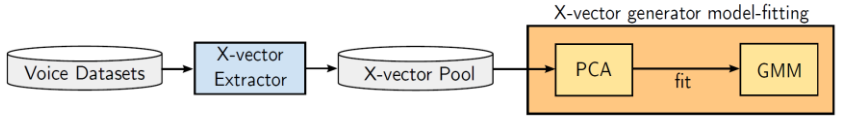
M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

O1: [Turner 2020]

- Keeping original distribution of cosine distances between speaker x-vectors
- GMM for sampling x-vectors in a PCA-reduced space with the following reconstruction of x-vectors of the original dimension



- ✗ Easy to identify system applied
- ✗ Voices lack the diversity of natural voices
- ✗ Application to same voice twice produces overly similar voices



*The figures are copied from the presentation [Turner 2020] Speaker Anonymization with Distribution-Preserving X-Vector Generation for the VoicePrivacy Challenge 2020. Henry Turner, Giulio Lovisotto, Ivan Martinovic

Approaches to x-vector anonymization: O1c1

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

O1c1: [Turner 2020]

- Forced dissimilarity between original and anonymized x-vectors

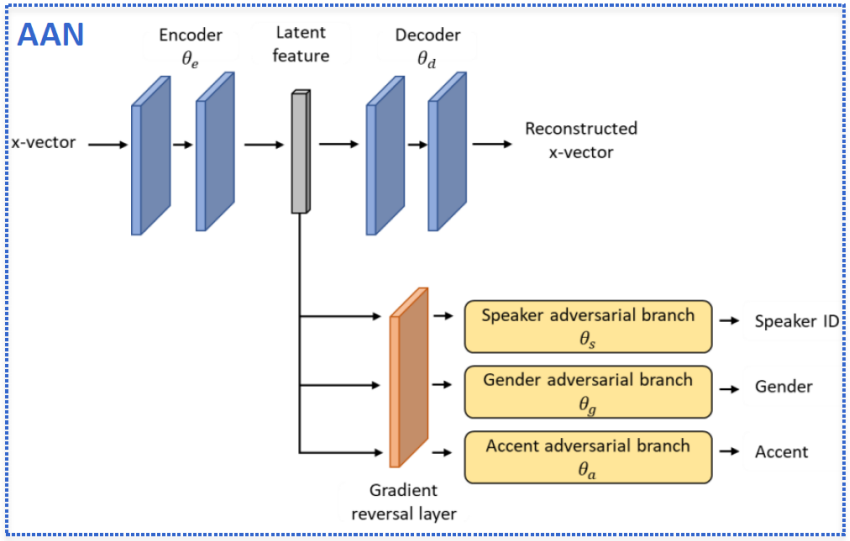
Approaches to x-vector anonymization: S2c1

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

S2c1: [Espinoza-Cuadros 2020]

- Domain-adversarial training
- Autoencoders using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction



*The figure is copied from the presentation [Espinoza-Cuadros 2020] Speaker De-identification System using Autoencoders and Adversarial Training. Fernando M. Espinoza-Cuadros, Juan M. Perero-Codosero, Javier Anton-Martın, Luis A. Hernandez-Gomez

Approaches to x-vector anonymization: S2

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	**B1**: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

S2: [Espinoza-Cuadros 2020]

- Domain-adversarial training
- Autoencoders using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction – **applied on the top of the B1 x-vector anonymization**

Approaches to x-vector anonymization: M1c2, M1c3, M1c4

System	Description
A2	**B1**: x-vector anonymization using singular value modification
A	B1: Different F0 extractors; x-vector anonymization using statistical-based ensemble regression modeling
★ O1	**B1:** x-vector anonymization keeping original distribution of cosine distances between speaker x-vectors; GMM for sampling vectors in a PCA-reduced space with the following reconstruction to the fake x-vectors of the original dimension
O1c1	**O1:** with forced dissimilarity between original and generated x-vectors
★ S2	**S2c1**: applied on the top of the B1 x-vector anonymization
S2c1	**B1:** x-vector anonymization using domain-adversarial training, autoencoders: using gender, accent, speaker id outputs corresponding to adversarial branches in ANN for x-vector reconstruction

M1c2	**B1:** copy-synthesis (original x-vectors)
M1c3	**B1**: x-vectors provided to SS AM are anonymized, x-vectors provided to NSF are original
M1c4	**B1**: x-vectors provided to SS AM are original, x-vectors provided to NSF are anonymized

[Champion 2020]

M1c2:

- Copy-synthesis (original x-vectors)

M1c3:

- X-vectors provided to SS AM are anonymized; x-vectors provided to NSF are original

M1c4:

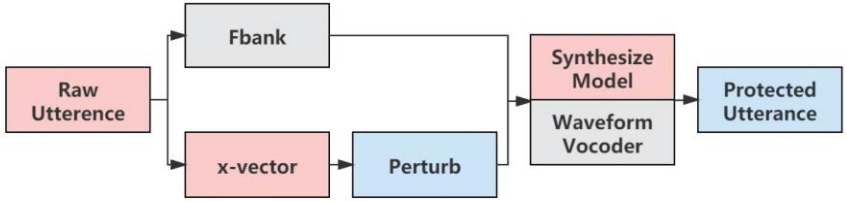
- X-vectors provided to SS AM are original; x-vectors provided to NSF are anonymized

Participants' systems: other approaches

★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm	Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius	
I1	modifications in formants, F0 and speaking rate	

Participants' systems: other approaches

- **K2:** [Han 2020]
  - Anonymization using x-vectors and SS models
  - Voice-indistinguishability metric
  - Speaker pool: test set itself

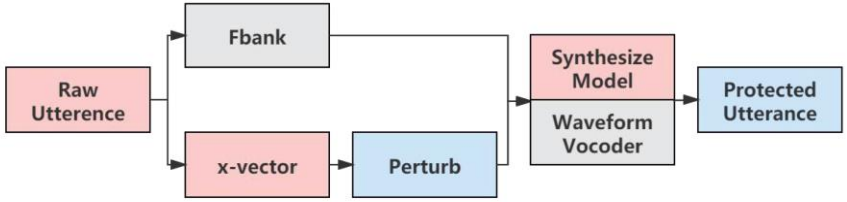


*The figure is copied from the presentation: [Han 2020] System Description for Voice Privacy Challenge. Yaowei Han, Sheng Li, Yang Cao, Masatoshi Yoshikawa

★ **K2**	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm	Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius	
I1	modifications in formants, F0 and speaking rate	

Participants' systems: other approaches

- **K2:** [Han 2020]
  - Anonymization using x-vectors and SS models
  - Voice-indistinguishability metric
  - Speaker pool: test set itself



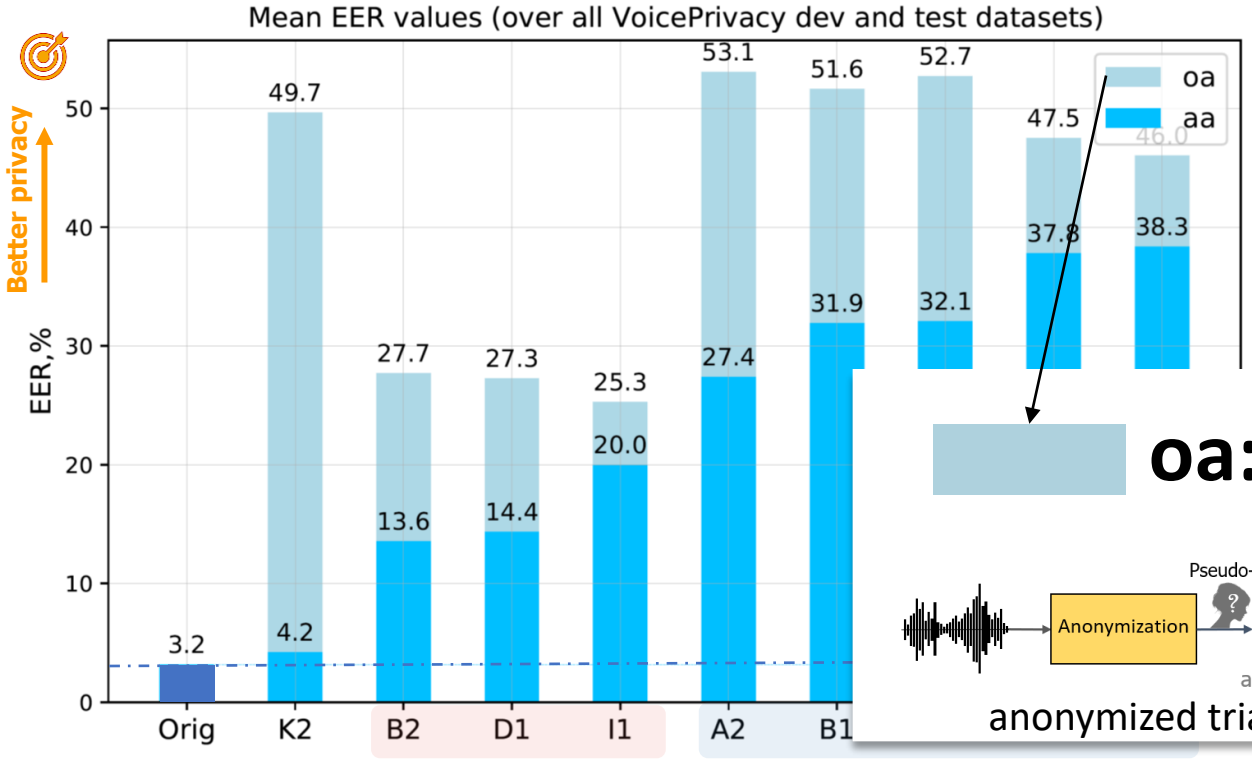
*The figure is copied from the presentation: [Han 2020] System Description for Voice Privacy Challenge. Yaowei Han, Sheng Li, Yang Cao, Masatoshi Yoshikawa

- **I1:** [Dubagunta 2020]
  - Modifications in formants, F0 and speaking rate

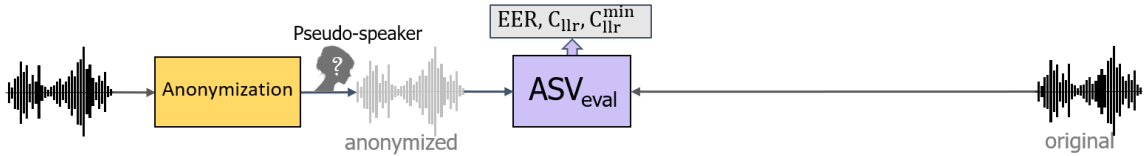
★ K2	anonymization using x-vectors and SS models: Voice-Indistinguishability metric; a waveform vocoder based on Griffin-Lim algorithm	Speaker pool: test set itself
D1	**B2**: additional modifications in pole radius	
I1	modifications in formants, F0 and speaking rate	

Objective evaluation results: EER

Results for selected primary systems

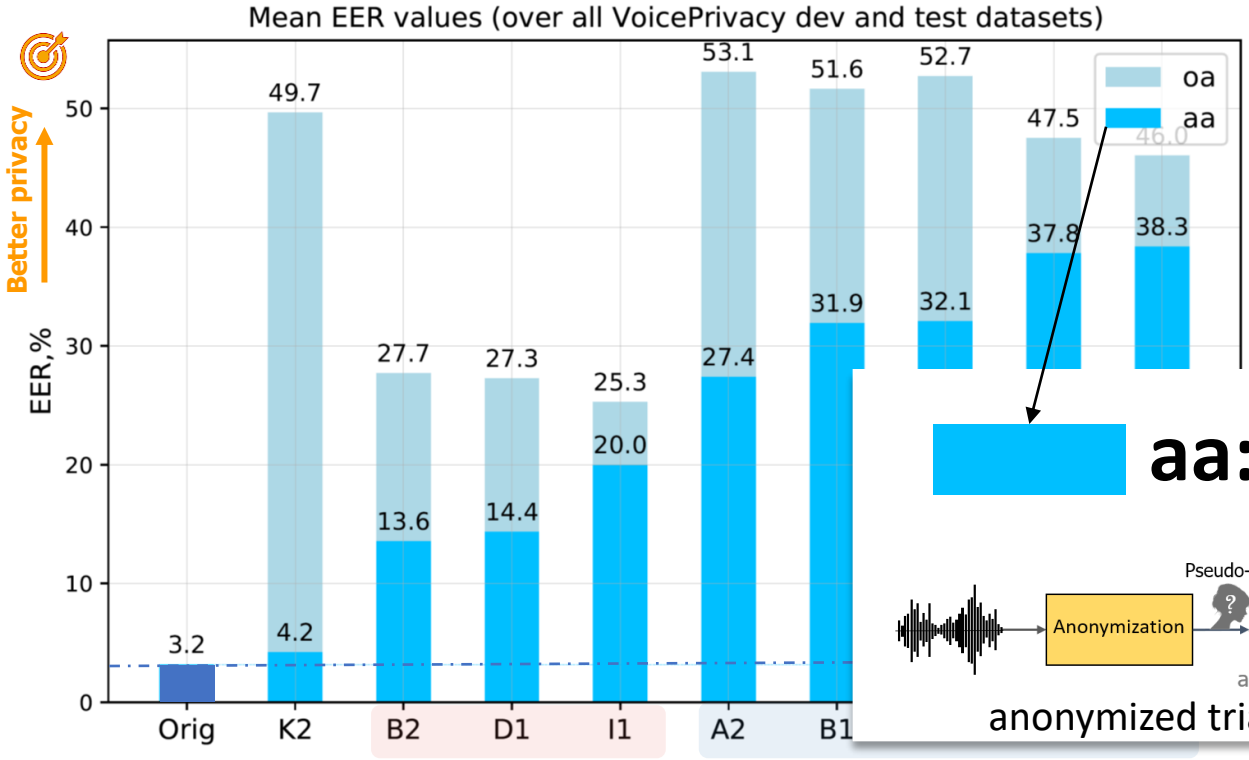


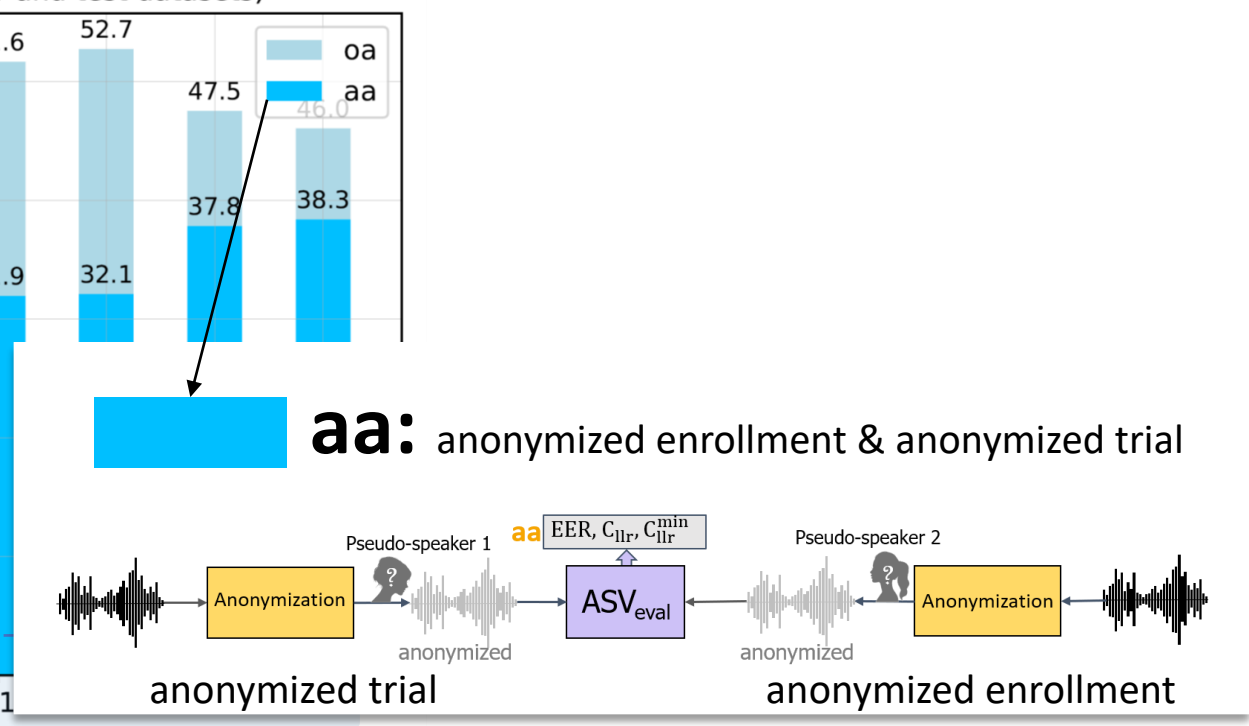
oa: original enrollment & anonymized trial



Objective evaluation results: EER

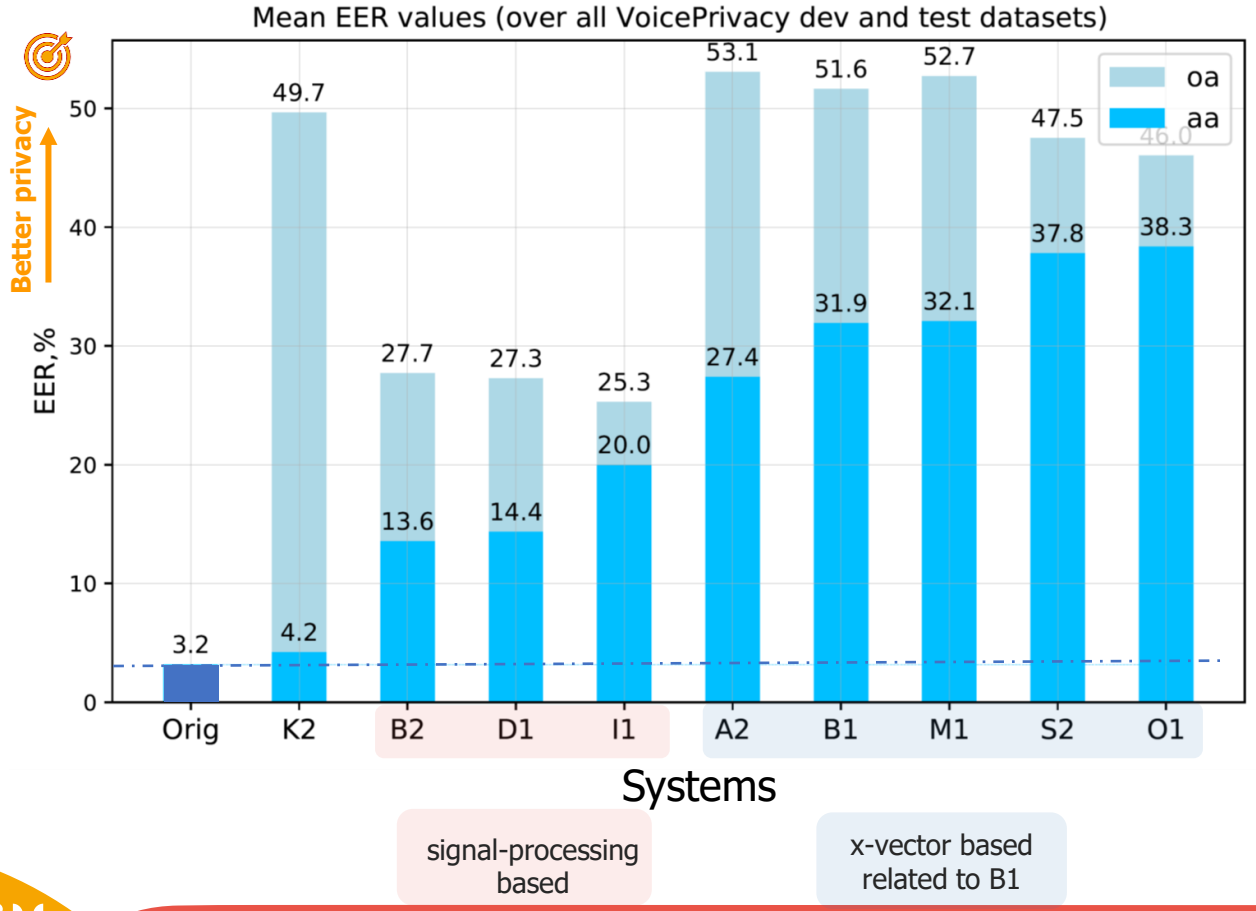
Results for selected primary systems





Objective evaluation results: EER

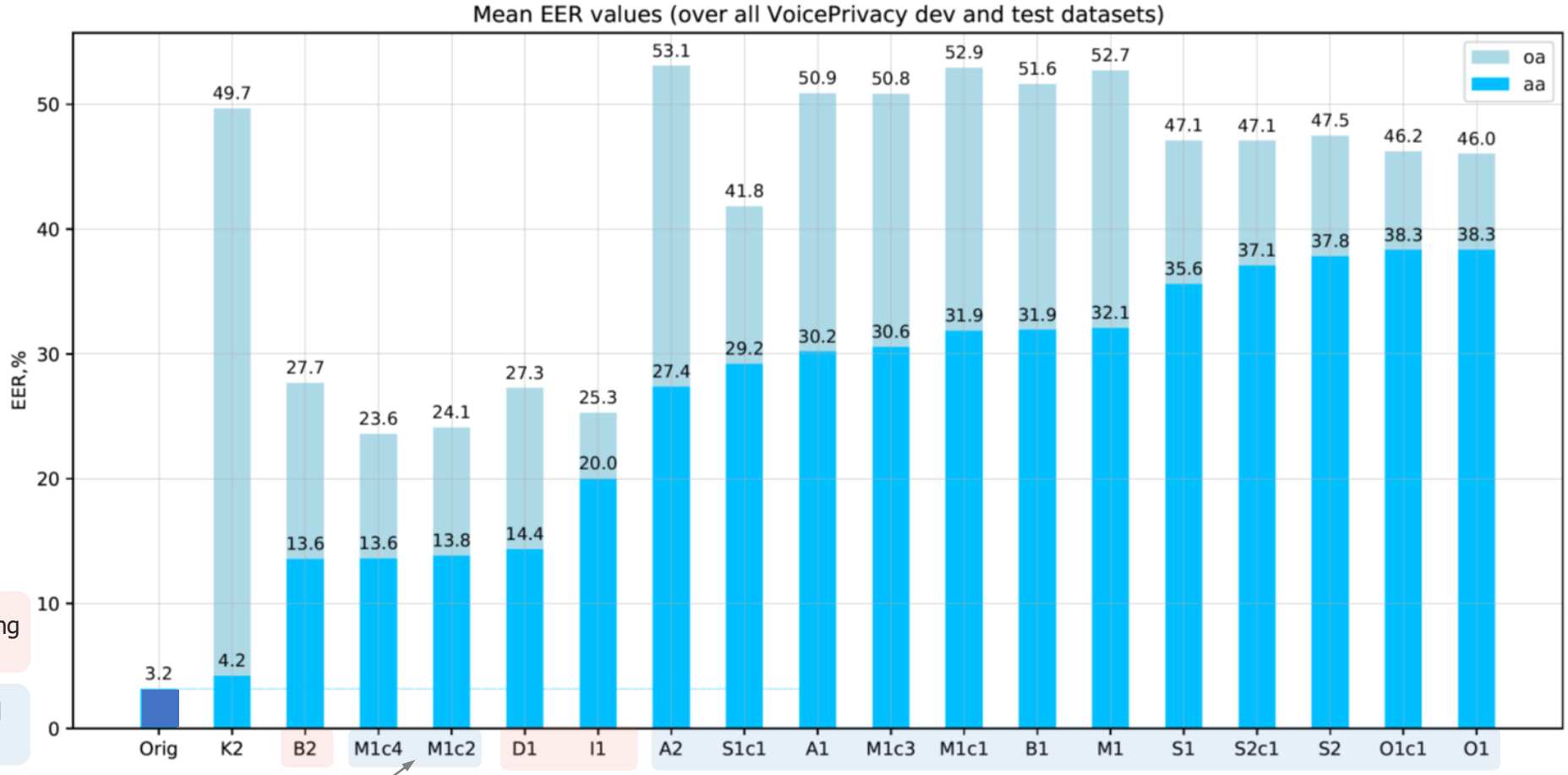
Results for selected primary systems



- **oa** – original enrollment, anonymized trial
- **aa** – anonymized enrollment, anonymized trial

- Anonymization of only the trial data greatly increases the EER (**oa**) => **anonymization effectively increases the users' privacy**
- Anonymization using pure **signal-processing** methods {**B2, D1, I1**} is **less effective** than methods related to B1
- Anonymized enrollment data result in a much lower EER (aa) for all the systems.

Objective evaluation results: EER for all systems



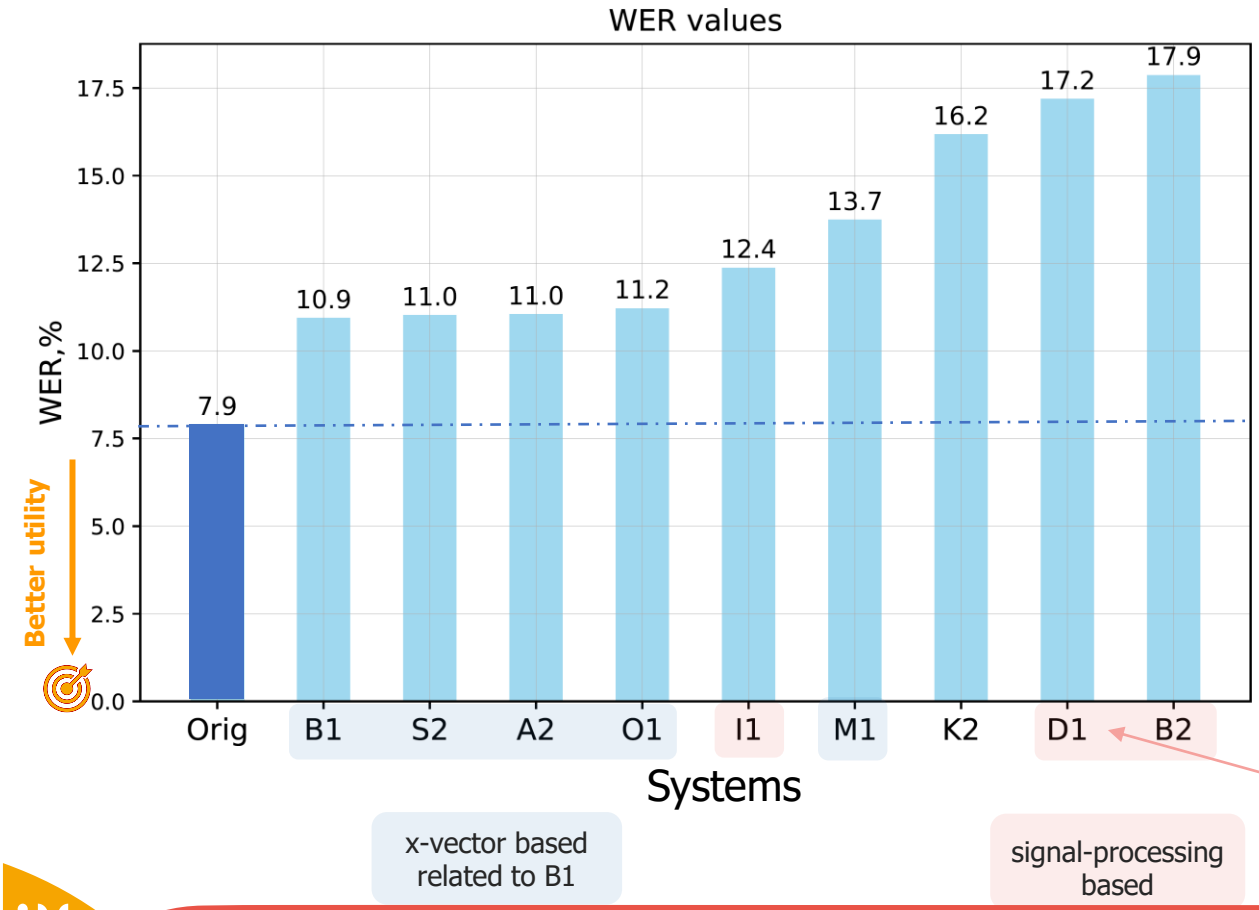
signal-processing based

x-vector based related to B1

- **M1c4**: x-vectors for SS AM original; for NSF – anonymized
- **M1c2**: copy-synthesis

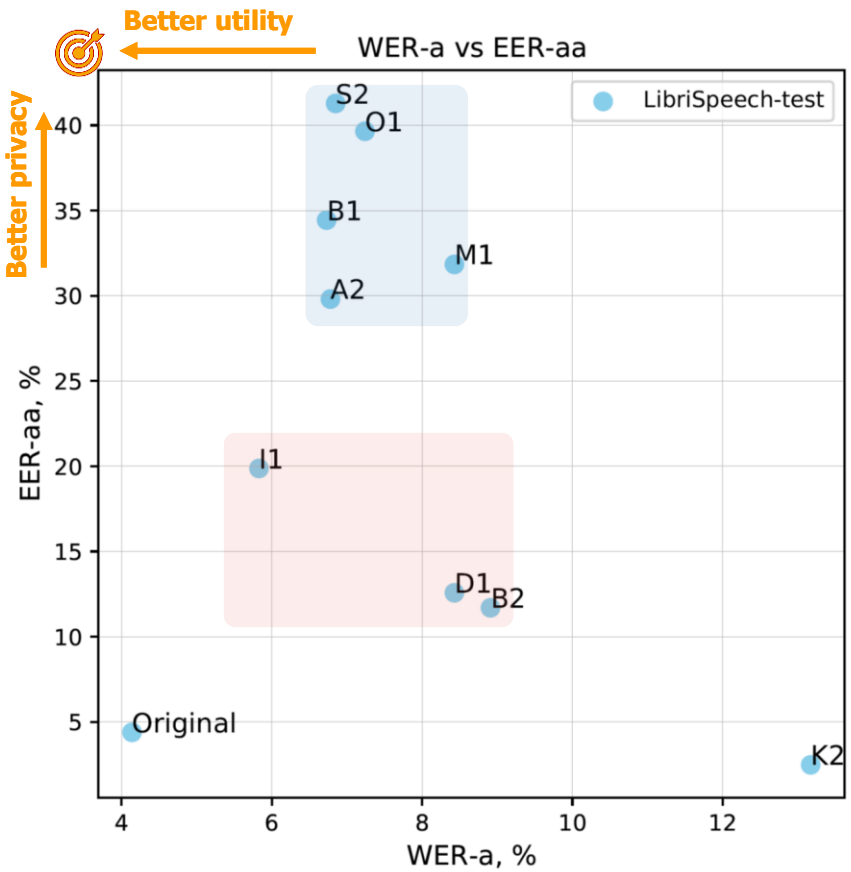
Objective evaluation results: WER

Results for selected primary systems



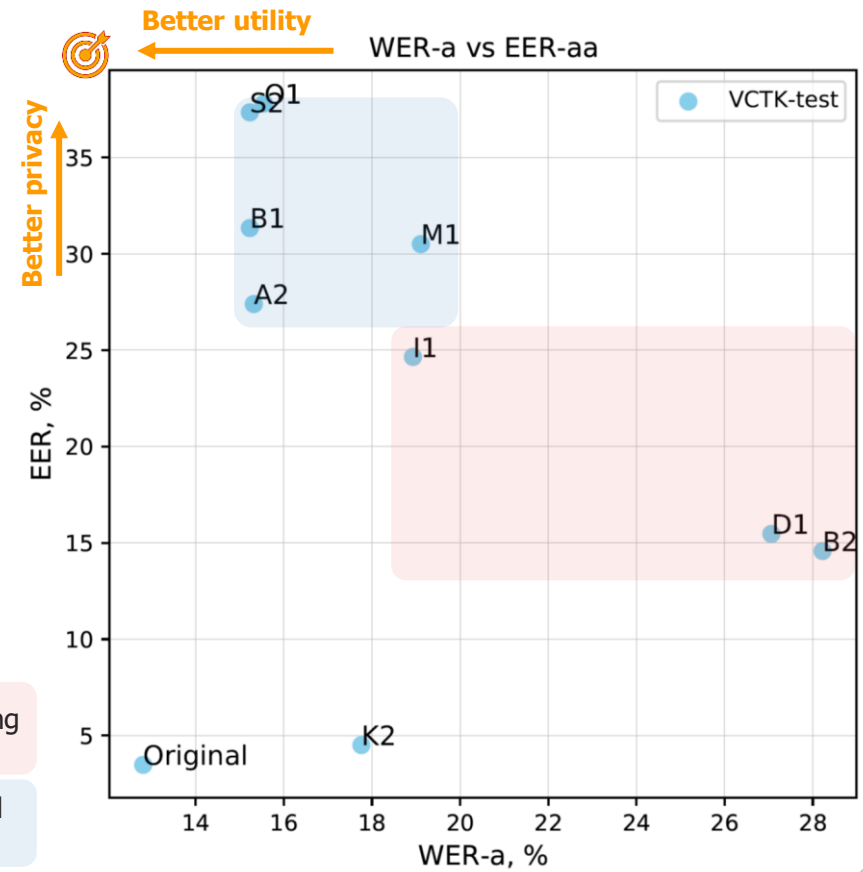
- **Anonymization incurs a large WER increase** for all the systems
- **Better** results for systems using **x-vector based anonymization related to B1**

Objective evaluation results: WER vs EER



signal-processing based

x-vector based related to B1



Participants' findings

- ✓ Participants proposed and investigated various anonymization approaches providing improvements in some test-cases/metrics over the baseline anonymization systems (B1, B2) including:
  - (B1) x-vector anonymization:
    - Distribution-preserving voice anonymization **O:** [Turner 2020]
    - Singular value modification and statistical-based decomposition with regression models **A:** [Mawalim 2020]
    - Domain-adversarial training and autoencoders **S:** [Espinoza-Cuadros 2020]
  - (B1) End-to-end ASR to extract BN features **M:** [Champion 2020]
  - (B2) Pole radius **D:** [Gupta 2020]
- ✓ Limitations of the baselines including:
  - (B1) Resynthesis by itself causes distortions in WER, increase in EER.
  - (B1) Not only x-vectors contain sensitive information, some leakage can be found in BNs, F0.
  - (B1) Anonymized x-vectors have different properties to original x-vectors **O:** [Turner 2020]

  A: [Mawalim 2020], **M:** [Champion 2020]
- ✓ Other anonymization approaches:
  - X-vector based anonymization using the voice-indistinguishability metric and SS models **K:** [Han 2020]
  - Signal-processing approach based on formant-shifting **I:** [Dubagunta 2020]

Participants' findings

- Participants proposed and investigated various anonymization approaches providing improvements in some test-cases/metrics over the baseline anonymization systems (B1, B2) including:
  - (B1) x-vector anonymization:
    - Distribution-preserving voice anonymization **O:** [Turner 2020]
    - Singular value modification and statistical-based decomposition with regression models **A:** [Mawalim 2020]
    - Domain-adversarial training and autoencoders **S:** [Espinoza-Cuadros 2020]
  - (B1) End-to-end ASR to extract BN features **M:** [Champion 2020]
  - (B2) Pole radius **D:** [Gupta 2020]
- Limitations of the baselines including:
  - (B1) Resynthesis by itself causes distortions in WER, increase in EER.
  - (B1) Not only x-vectors contain sensitive information, some leakage can be found in BNs, F0.
  - (B1) Anonymized x-vectors have different properties to original x-vectors **O:** [Turner 2020]

A: [Mawalim 2020],
M: [Champion 2020]

- Other anonymization approaches:
  - X-vector based anonymization using the voice-indistinguishability metric and SS models **K:** [Han 2020]
  - Signal-processing approach based on formant-shifting **I:** [Dubagunta 2020]

Participants' findings

- Participants proposed and investigated various anonymization approaches providing improvements in some test-cases/metrics over the baseline anonymization systems (B1, B2) including:
  - (B1) x-vector anonymization:
    - Distribution-preserving voice anonymization **O:** [Turner 2020]
    - Singular value modification and statistical-based decomposition with regression models **A:** [Mawalim 2020]
    - Domain-adversarial training and autoencoders **S:** [Espinoza-Cuadros 2020]
  - (B1) End-to-end ASR to extract BN features **M:** [Champion 2020]
  - (B2) Pole radius **D:** [Gupta 2020]
- Limitations of the baselines including:
  - (B1) Resynthesis by itself causes distortions in WER, increase in EER.
  - (B1) Not only x-vectors contain sensitive information, some leakage can be found in BNs, F0.
  - (B1) Anonymized x-vectors have different properties to original x-vectors **O:** [Turner 2020]

A: [Mawalim 2020],
M: [Champion 2020]

- Other anonymization approaches:
  - X-vector based anonymization using the voice-indistinguishability metric and SS models **K:** [Han 2020]
  - Signal-processing approach based on formant-shifting **I:** [Dubagunta 2020]

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)

- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions

- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.

- Potential for improvement

- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)

- The considered metrics do not evaluate all the requirements for anonymization

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)
- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions
- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.
- Potential for improvement
- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)
- The considered metrics do not evaluate all the requirements for anonymization

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)
- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions
- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.
- Potential for improvement
- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)
- The considered metrics do not evaluate all the requirements for anonymization

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)
- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions
- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.
- Potential for improvement
- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)
- The considered metrics do not evaluate all the requirements for anonymization

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)
- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions
- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.
- Potential for improvement
- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)
- The considered metrics do not evaluate all the requirements for anonymization

Conclusions

- **2** classes of anonymization methods:
  - x-vectors-based with speech synthesis models (B1 and related methods)
  - signal-processing based (B2 and others)
- Anonymization using **x-vector-based anonymization techniques related to B1** in average is **more effective** than signal-based processing techniques: better privacy (EER) and utility (WER) but there are some exceptions
- Systems perform and are ranked differently for different attack models. There is no system which is the best for all metrics, all datasets and attack models.
- Potential for improvement
- Investigate other attack models and downstream tasks – in post-evaluation (the following part of this presentation)
- The considered metrics do not evaluate all the requirements for anonymization

References: participants' papers

- **A:** [Mawalim 2020] X-Vector Singular Value Modification and Statistical-Based Decomposition with Ensemble Regression Modeling for Speaker Anonymization System. Candy Olivia Mawalim, Kasorn Galajit, Jessada Karnjana, Masashi Unoki
- **D:** [Gupta 2020] Design of Voice Privacy System using Linear Prediction. Priyanka Gupta, Gauri P. Prajapati, Shrishti Singh, Madhu R. Kamble, Hemant A. Patil
- **I:** [Dubagunta 2020] Adjustable Deterministic Pseudonymisation of Speech: Idiap-NKI's submission to VoicePrivacy 2020 Challenge. S. Pavankumar Dubagunta, Rob J.J.H. van Son and Mathew Magimai.-Doss
- **K:** [Han 2020] System Description for Voice Privacy Challenge. Yaowei Han, Sheng Li, Yang Cao, Masatoshi Yoshikawa
- **M:** [Champion 2020] Speaker information modification in the VoicePrivacy 2020 toolchain. Pierre Champion, Denis Jouvet, Anthony Larcher.
- **O:** [Turner 2020] Speaker Anonymization with Distribution-Preserving X-Vector Generation for the VoicePrivacy Challenge 2020. Henry Turner, Giulio Lovisotto, Ivan Martinovic
- **S:** [Espinoza-Cuadros 2020] Speaker De-identification System using Autoencoders and Adversarial Training. Fernando M. Espinoza-Cuadros, Juan M. Perero-Codosero, Javier Anton-Martın, Luis A. Hernandez-Gomez
- [Chien-Lin Huang 2020] Analysis of PingAn Submission in the VoicePrivacy 2020 Challenge. Chien-Lin Huang

References VoicePrivacy challenge

- VoicePrivacy site: https://www.voiceprivacychallenge.org/
- Baseline software: https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2020
- Evaluation plan: https://www.voiceprivacychallenge.org/docs/VoicePrivacy_2020_Eval_Plan_v1_3.pdf

- [Tomashenko 2020] Introducing the VoicePrivacy initiative. Natalia Tomashenko, Brij Mohan Lal Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, Massimiliano Todisco

Alternative anonymization metrics:

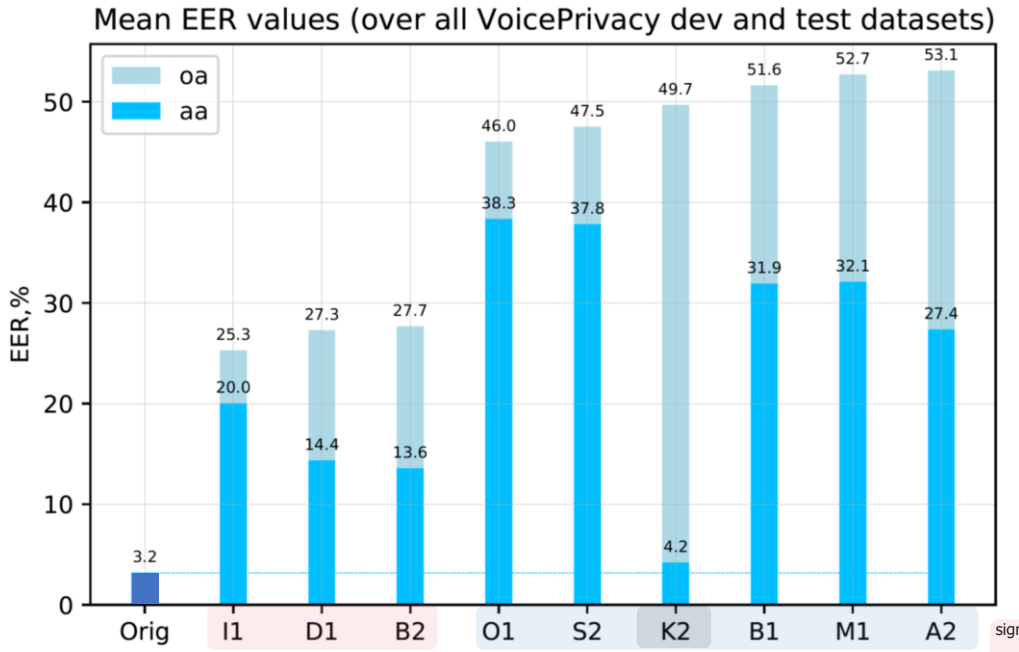
- [Noe 2020] Speech Pseudonymisation Assessment Using Voice Similarity Matrices. Paul-Gauthier Noe, Jean-Francois Bonastre, Driss Matrouf, Natalia Tomashenko, Andreas Nautsch and Nicholas Evans
- [Nautsch 2020] The Privacy ZEBRA: Zero Evidence Biometric Recognition Assessment. Andreas Nautsch, Jose Patino, Natalia Tomashenko, Junichi Yamagishi, Paul-Gauthier Noe, Jean-Francois Bonastre, Massimiliano Todisco, Nicholas Evans
- [Maouche 2020] A comparative study of speech anonymization metrics. Mohamed Maouche, Brij Mohan Lal Srivastava, Nathalie Vauquier, Aurélien Bellet, Marc Tommasi, Emmanuel Vincent

References

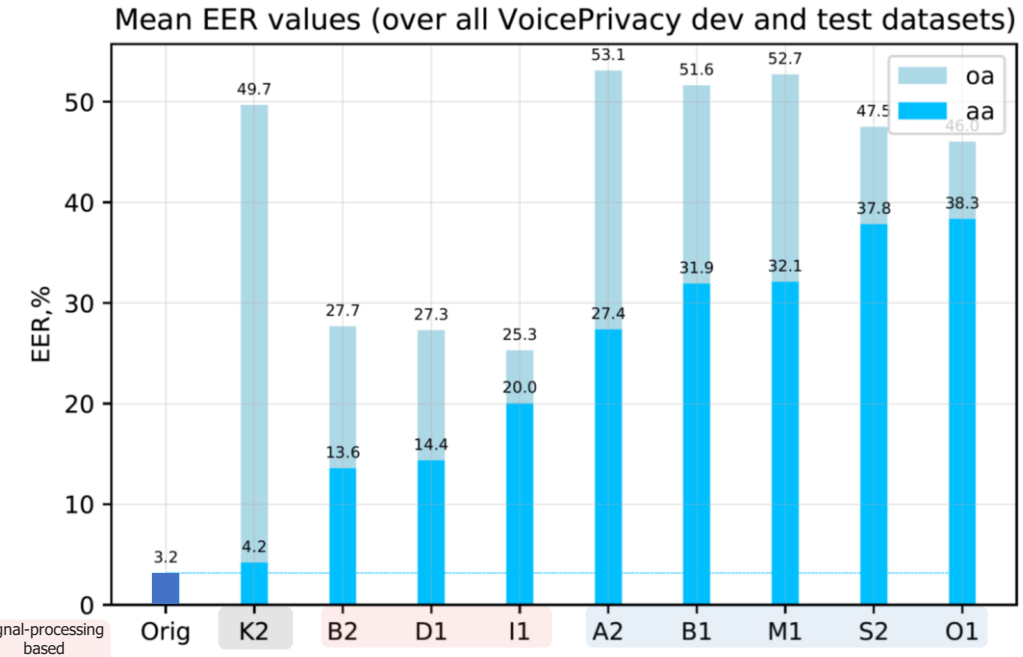
- [Cohen-Hadria 2019] Voice Anonymization in Urban Sound Recordings. A. Cohen-Hadria, M. Cartwright, B. McFee, & J.P. Bello
- [Gontier 2020] Privacy aware acoustic scene synthesis using deep spectral feature inversion. F. Gontier, M. Lagrange, C. Lavandier, & J.-F. Petiot
- [Pathak 2013] Privacy preserving speech processing: cryptographic and string-matching frameworks show promise. M. A. Pathak, B. Raj, S. D. Rane, & P. Smaragdis.
- [Smaragdis 2007] A framework for secure speech recognition. P. Smaragdis and M. Shashanka.
- [Leroy 2019] Federated learning for keyword spotting. D. Leroy, A. Coucke, T. Lavril, T. Gisselbrecht, & J. Dureau
- [Hashimoto 2016] Privacy-preserving sound to degrade automatic speaker verification performance. K. Hashimoto, J. Yamagishi, and I. Echizen
- [Qian 2017] Voicemask: Anonymize and sanitize voice input on mobile devices. J. Qian, H. Du, J. Hou, L. Chen, T. Jung, X.-Y. Li, Y. Wang, & Y. Deng
- [Jin 2009] Voice convergin: Speaker de-identification by voice transformation. Q. Jin, A. Toth, T. Schultz, & A. Black
- [Fang 2019] Speaker anonymization using x-vector and neural waveform models. F. Fang, X.Wang, J. Yamagishi, I. Echizen, M. Todisco, N. Evans, & J.-F. Bonastre
- [Srivastava 2019] Privacy-preserving adversarial representation learning in ASR. B. M. L. Srivastava, A. Bellet, M. Tommasi, & E. Vincent
- [McAdams 1984] Spectral fusion, spectral parsing and the formation of the auditory image. S. McAdams
- [Patino 2020] Speaker anonymisation using the McAdams coefficient. J. Patino, M. Todisco, A. Nautsch, & N. Evans

Other results

Objective evaluation results: EER for primary systems



Sorted by **EER-oa**



Sorted by **EER-aa**

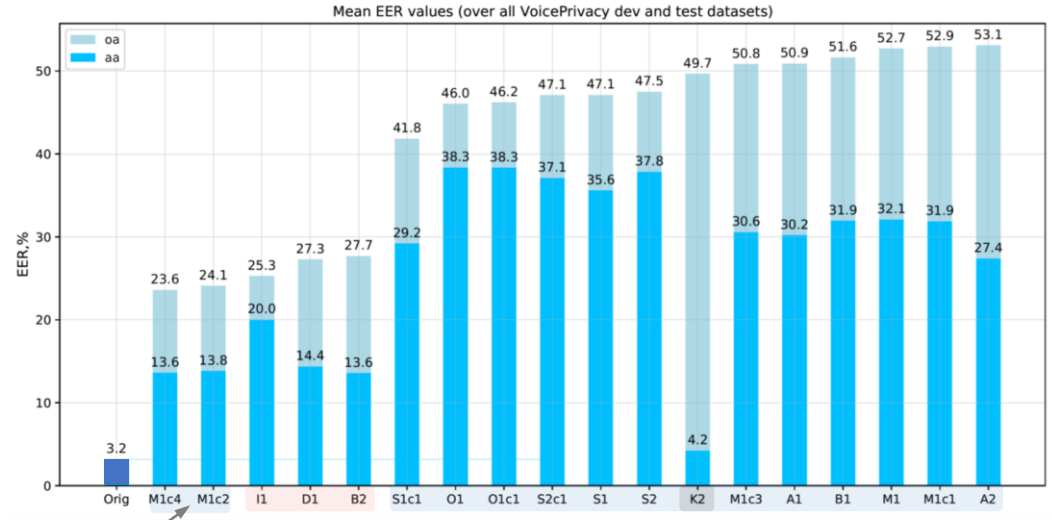
signal-processing based transformations

x-vector based

- Anonymization of **only the trial** data greatly increases the EER (oa) for x-vector based anonymization methods: **A2**, **M1**, **B1**, **K2**, **S2**, **O1** => anonymization effectively increases the users' privacy. Full anonymization (EER>50%) wrt to this attack model (oa) is achieved for **A2**, **M1**, **B1**.
- Anonymization using pure **signal-processing** methods (**B2**, **D1**, **I1**) is **less effective**

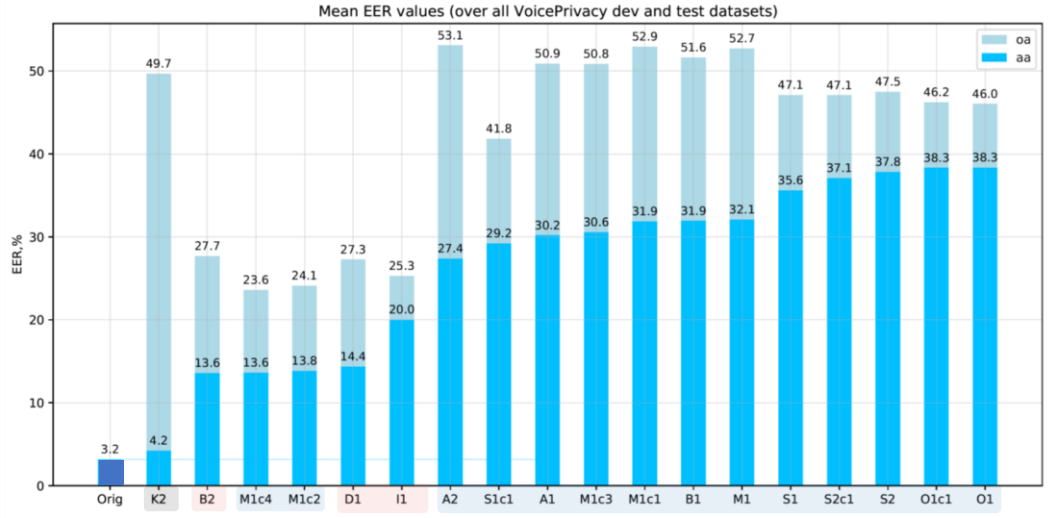
- Anonymized enrollment data result in a much lower EER (aa) for all the systems.
- The order of the system is different for EER-aa and EER-oa, but in all cases **anonymization is more effective** for **x-vector based** methods. Exception: **K2**.

Objective evaluation results: EER for all systems



- **M1c4**: x-vectors for SS AM original; for NSF – anonymized
- **M1c2**: copy-synthesis

Sorted by **EER-oa**

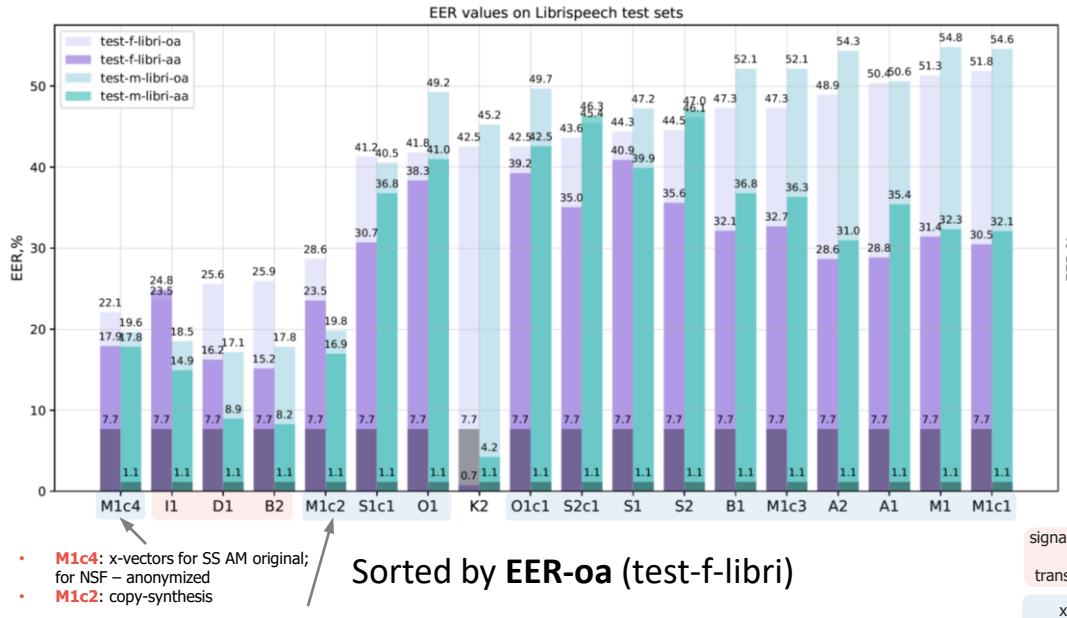


Sorted by **EER-aa**

signal-processing based transformations

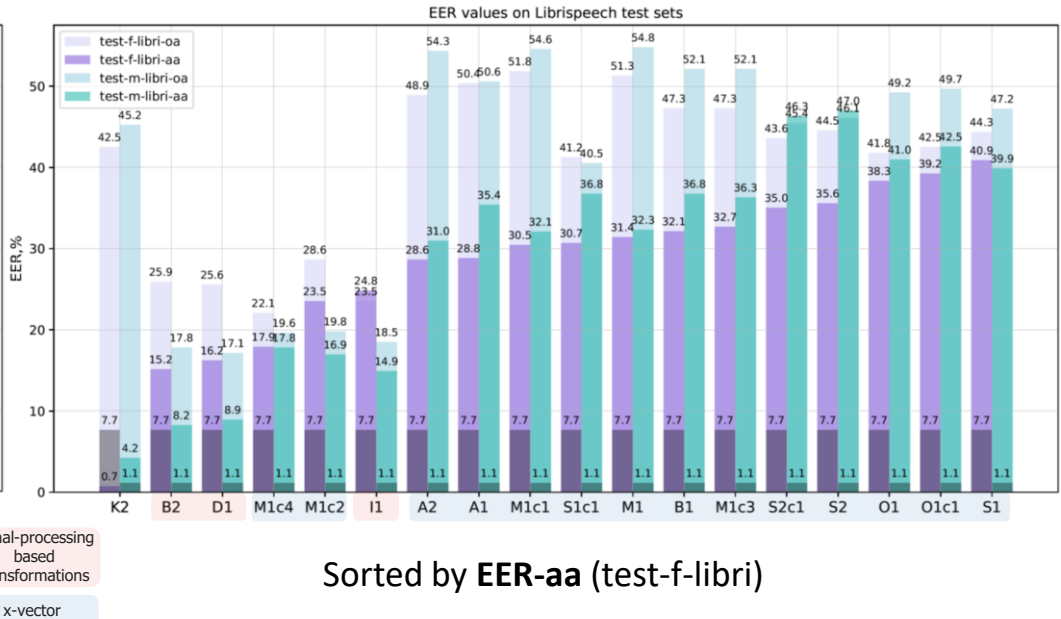
x-vector based

Objective evaluation results: EER on LibriSpeech test



- **M1c4**: x-vectors for SS AM original; for NSF – anonymized
- **M1c2**: copy-synthesis

Sorted by **EER-oa** (test-f-libri)



Sorted by **EER-aa** (test-f-libri)

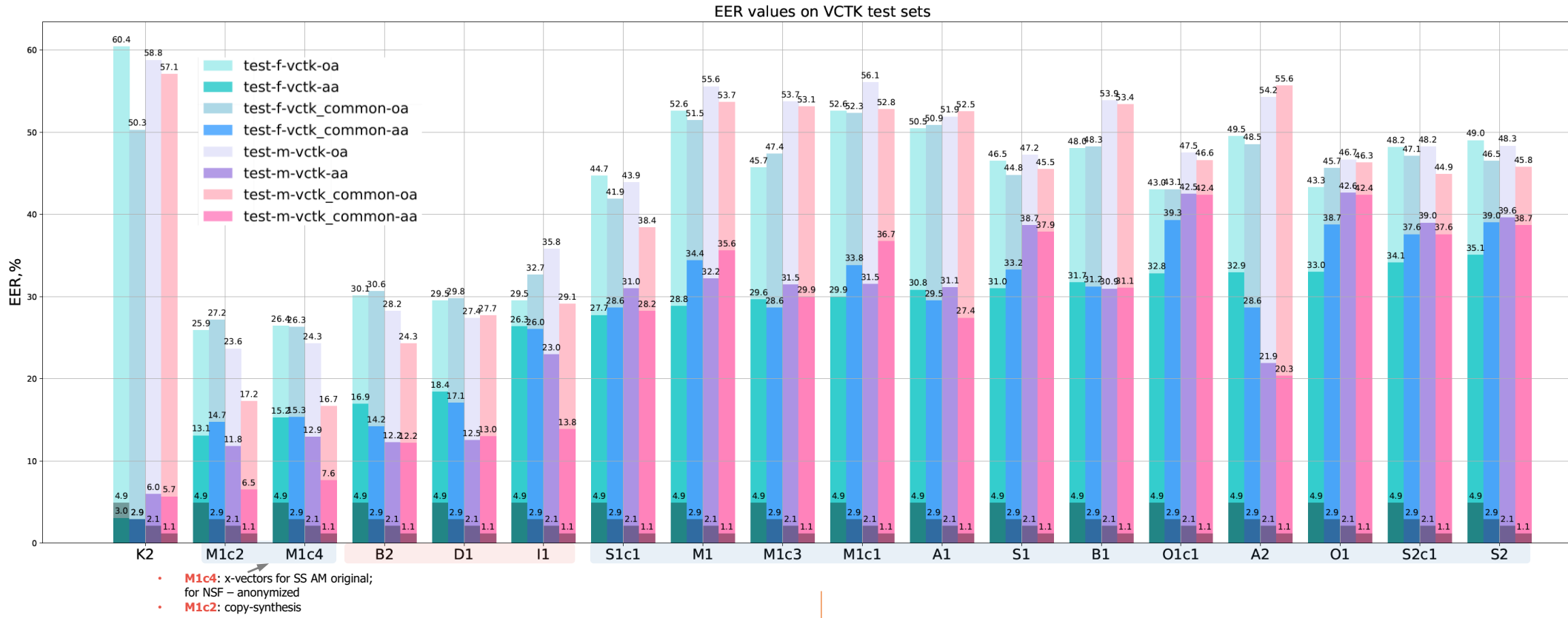
signal-processing based transformations

x-vector based

- **Baseline EER** (on original data) for **female** is much **higher** than for male speakers.
- **EER-oa** for **female** is
  - **lower** than for **male** for most of the **x-vector based** anonymization systems (exception: **M1c2**, **S1c1**, **M1c4**)
  - **higher** than for **male** for all **signal-processing based** methods.

- **EER-aa** for **female** is
  - lower than for male for most of the x-vector based anonymization systems (exception: **S1**, **M1c1**, **M1c4**);
  - and higher for all signal-processing based methods.
- For **K2**: **EER-aa** is significantly **decreased** for female in comparison with the baseline EER (while for EER-oa a high level anonymization is achieved by this system).

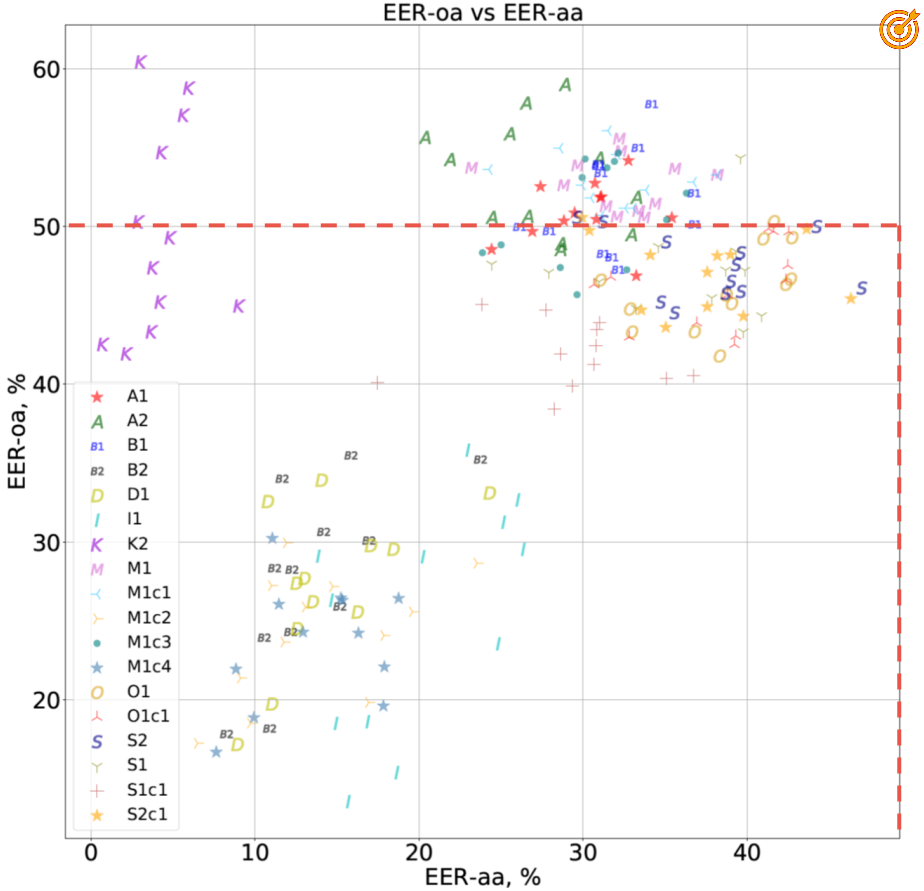
Objective evaluation results: EER on VCTK-test



- **Baseline EER** (on original data) for **female** is **higher** than for male speakers.
- **EER-aa** for **female** is **higher** than for **male** for all **signal-processing based** methods, but other methods perform differently

- For **K2**: **EER-aa** is **decreased** for female in comparison with the original EER (while for EER-oa for this dataset it has the highest level of anonymization among the systems).

Objective evaluation results: EER for different attack models

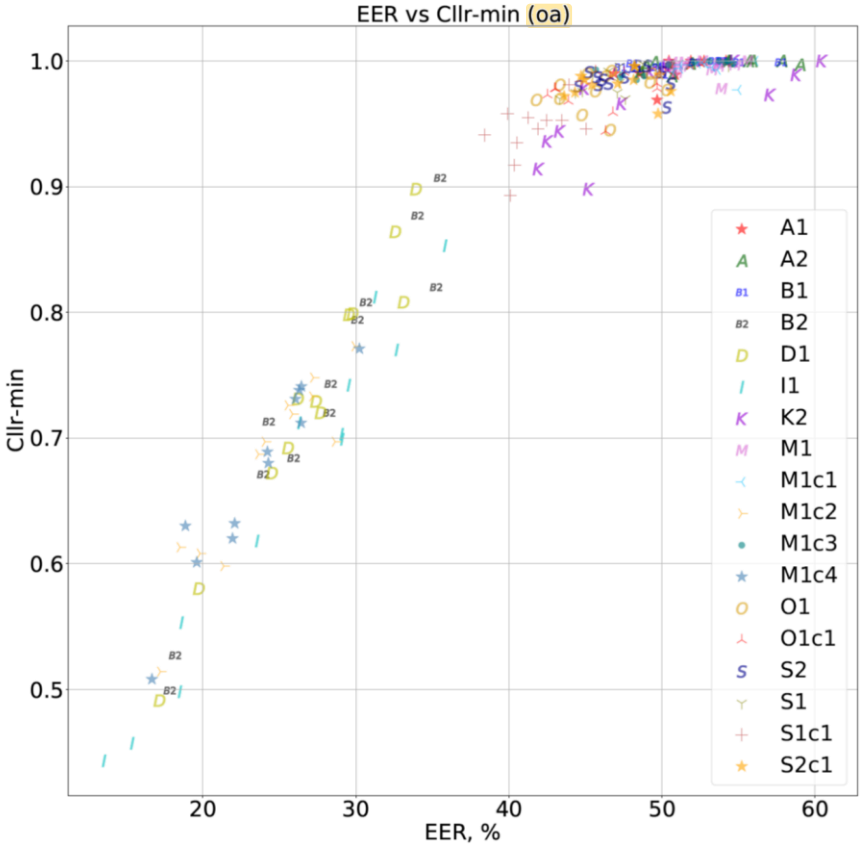


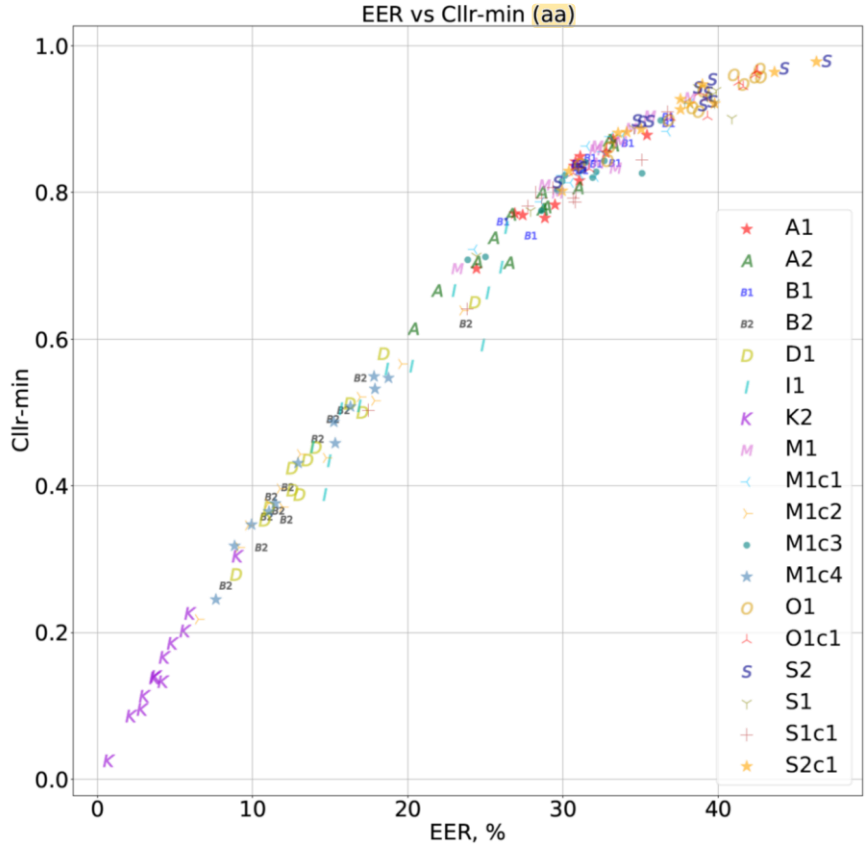
Each point corresponds to EER results for a particular dev or test dataset from the set of all 12 VoicePrivacy dev/test datasets.

- Many systems (especially {**A2, M1, B1**} and **K2** for some datasets), anonymization of the trial data greatly increases the EER (**oa**).
- For (**oa**) many systems have ERR > 50% => anonymization has been achieved ?
- Using anonymized enrollment data (**aa**) results in a much lower EER for most of the systems.
- Different performance for (**oa**) and (**aa)** evaluation cases => the choice of the optimal anonymization algorithm will depend an the attack model and available data.
- It is important to correctly define the attack model to avoid overestimated sense of privacy.

(**oa**) original enrollment, anonymized trials
(**aa**) anonymized enrollment, anonymized trials

Objective evaluation results: EER vs C_{llr}^{min}

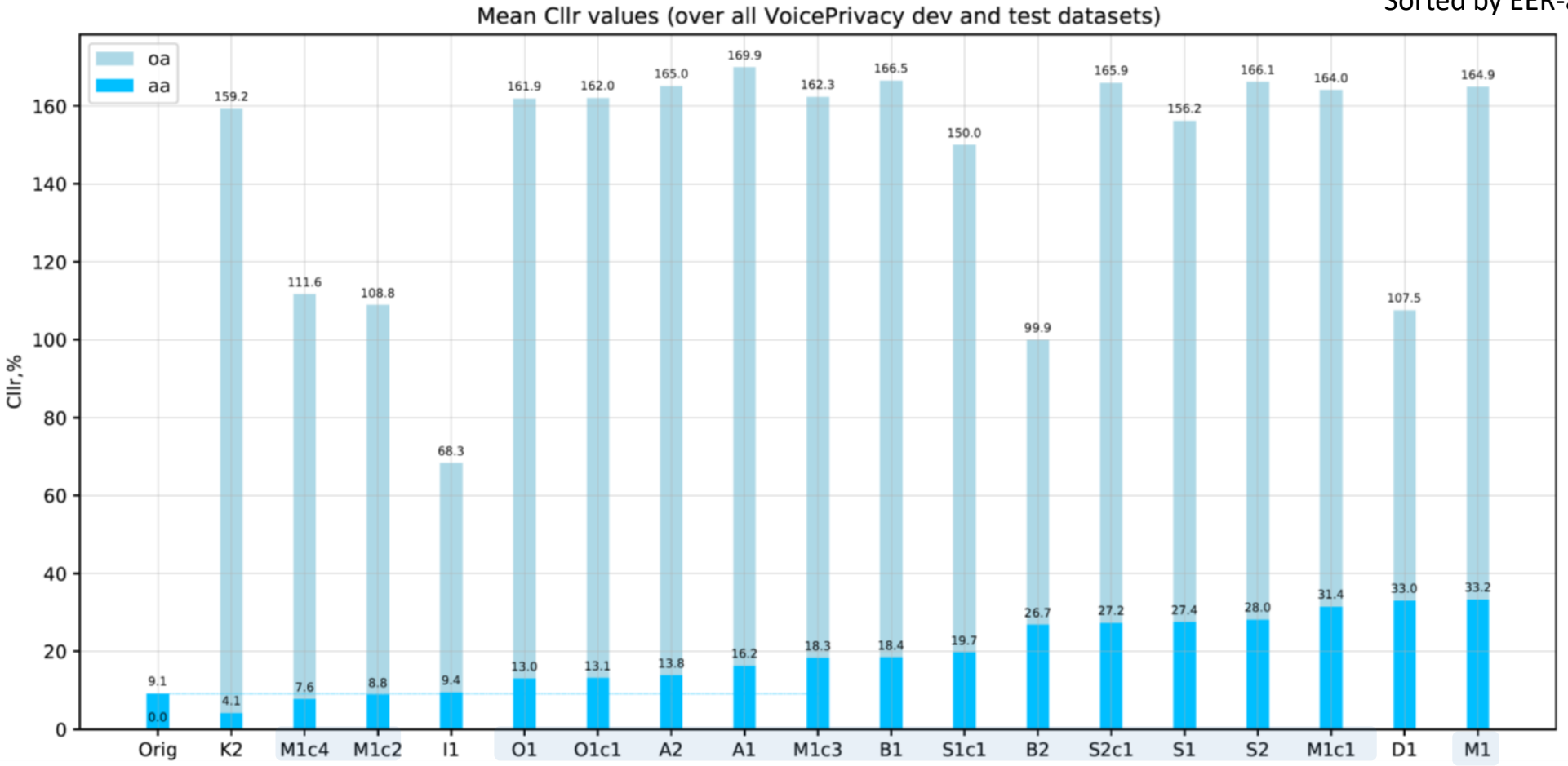




Each point corresponds to ASV_eval results for a particular dev or test dataset from the set of all 12 VoicePrivacy dev/test datasets.

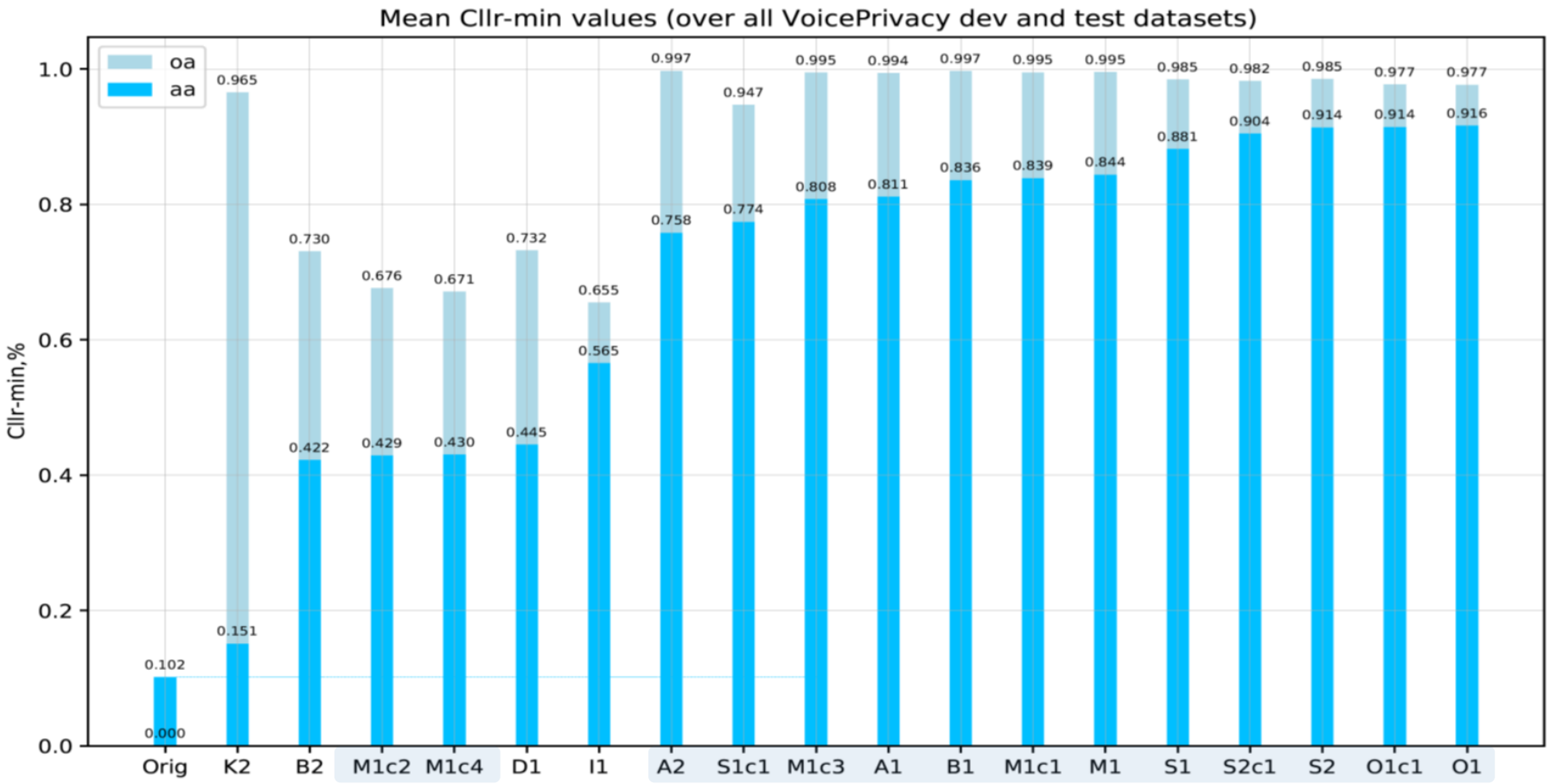
Objective evaluation results: C_{llr} all systems

Sorted by EER-aa

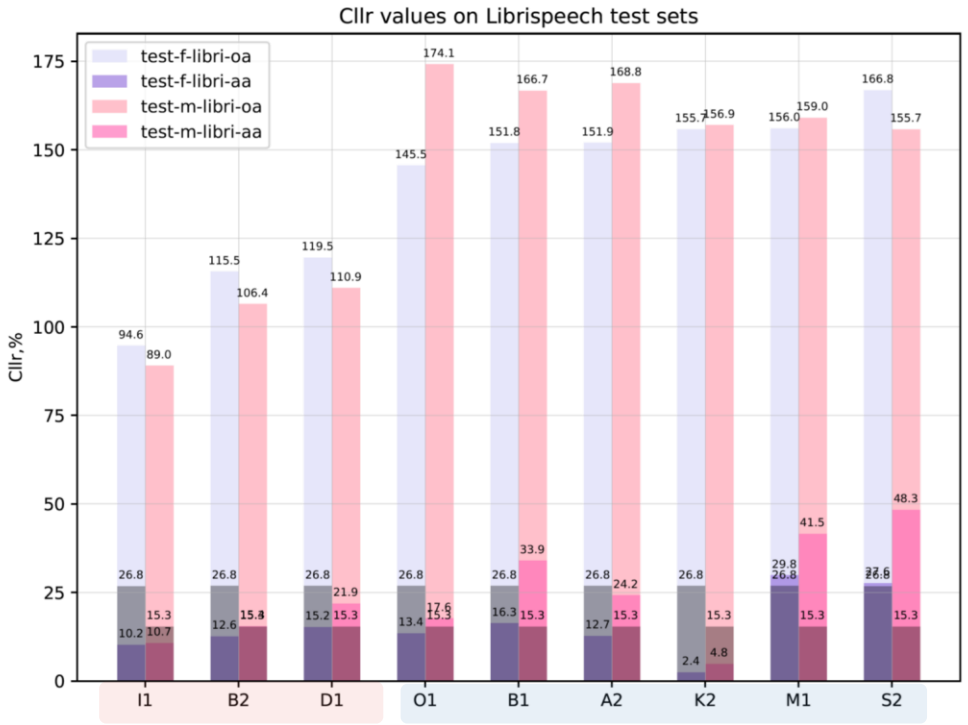


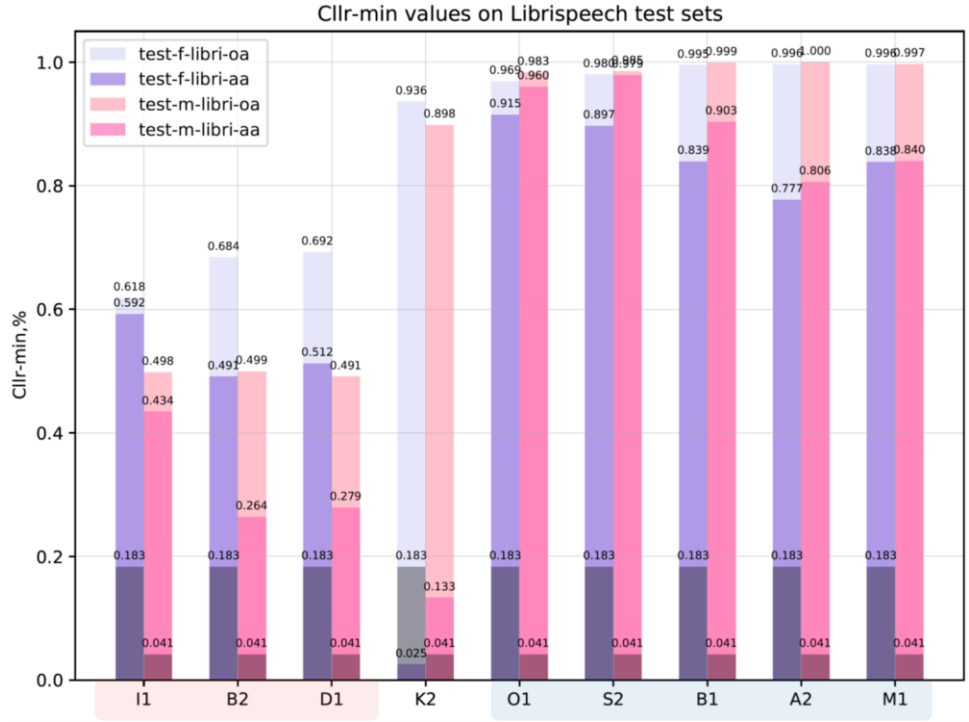
Objective evaluation results: C_{llr}^{min} all systems

Sorted by EER-aa



Objective evaluation results: C_{llr} & C_{llr}^{min} on Librispeech-test

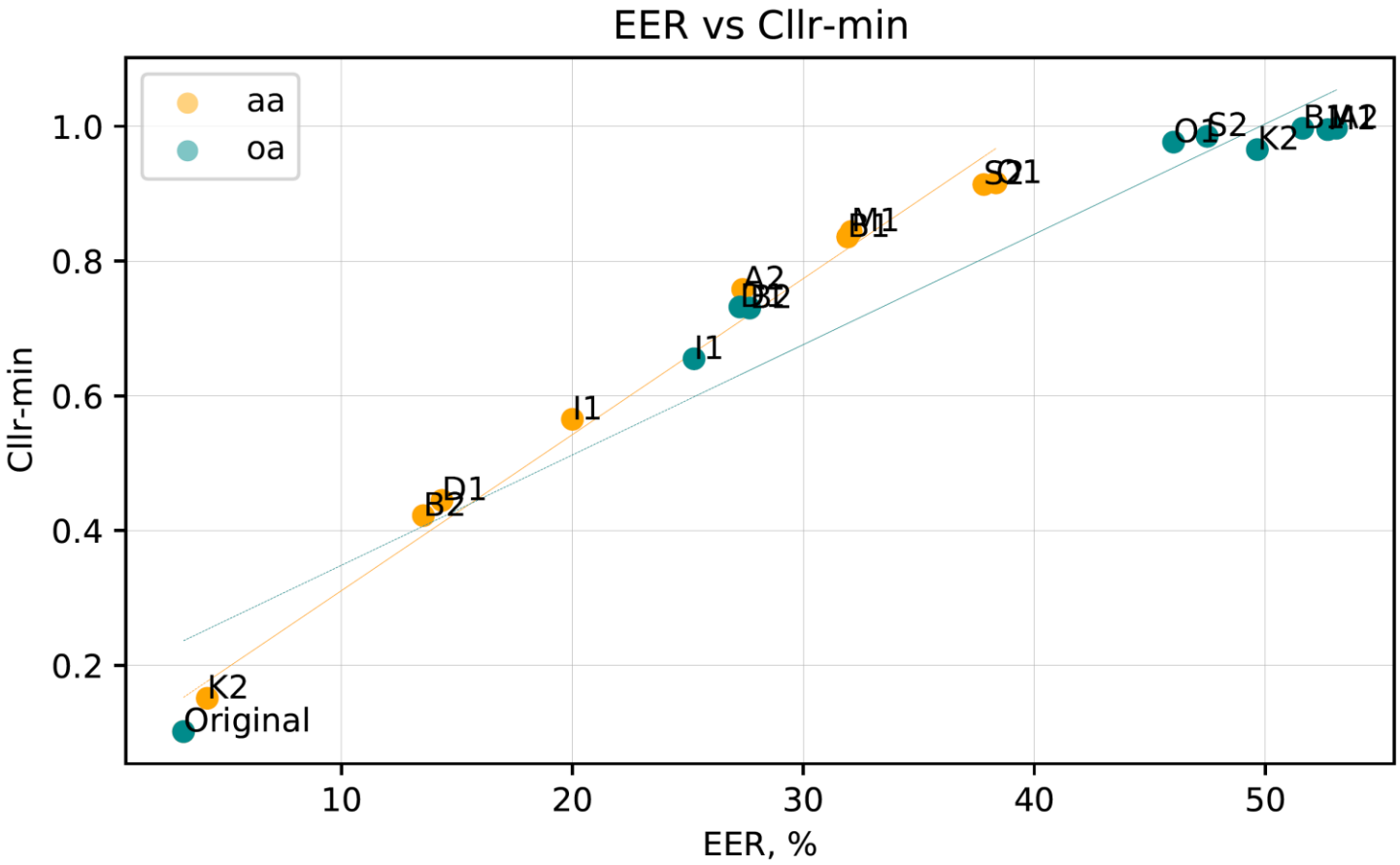




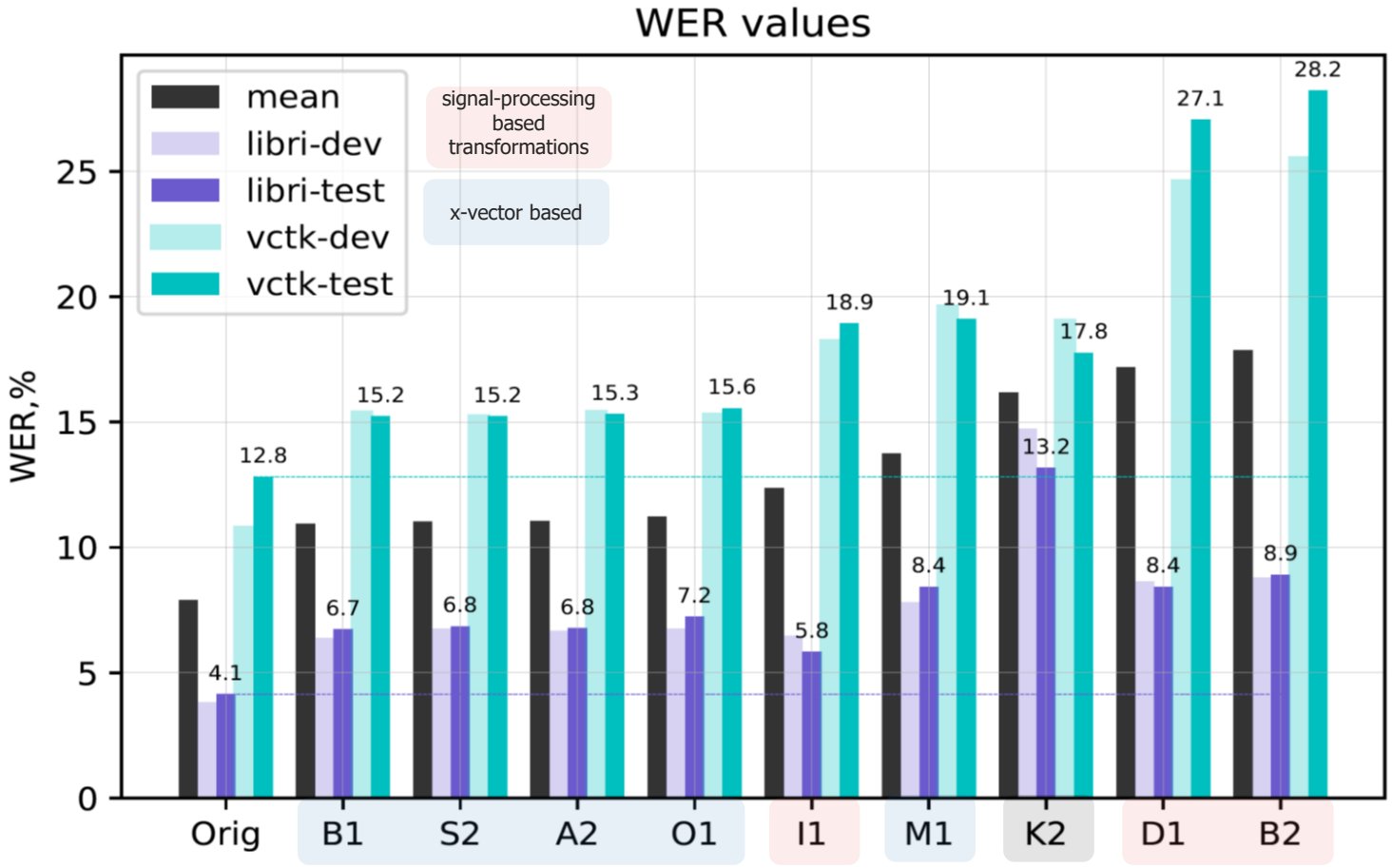
Sorted by oa f-test-libri

Objective evaluation results: C_{llr} & C_{llr}^{min} on VCTK-test

Objective evaluation results: mean EER vs C_{llr}^{min}



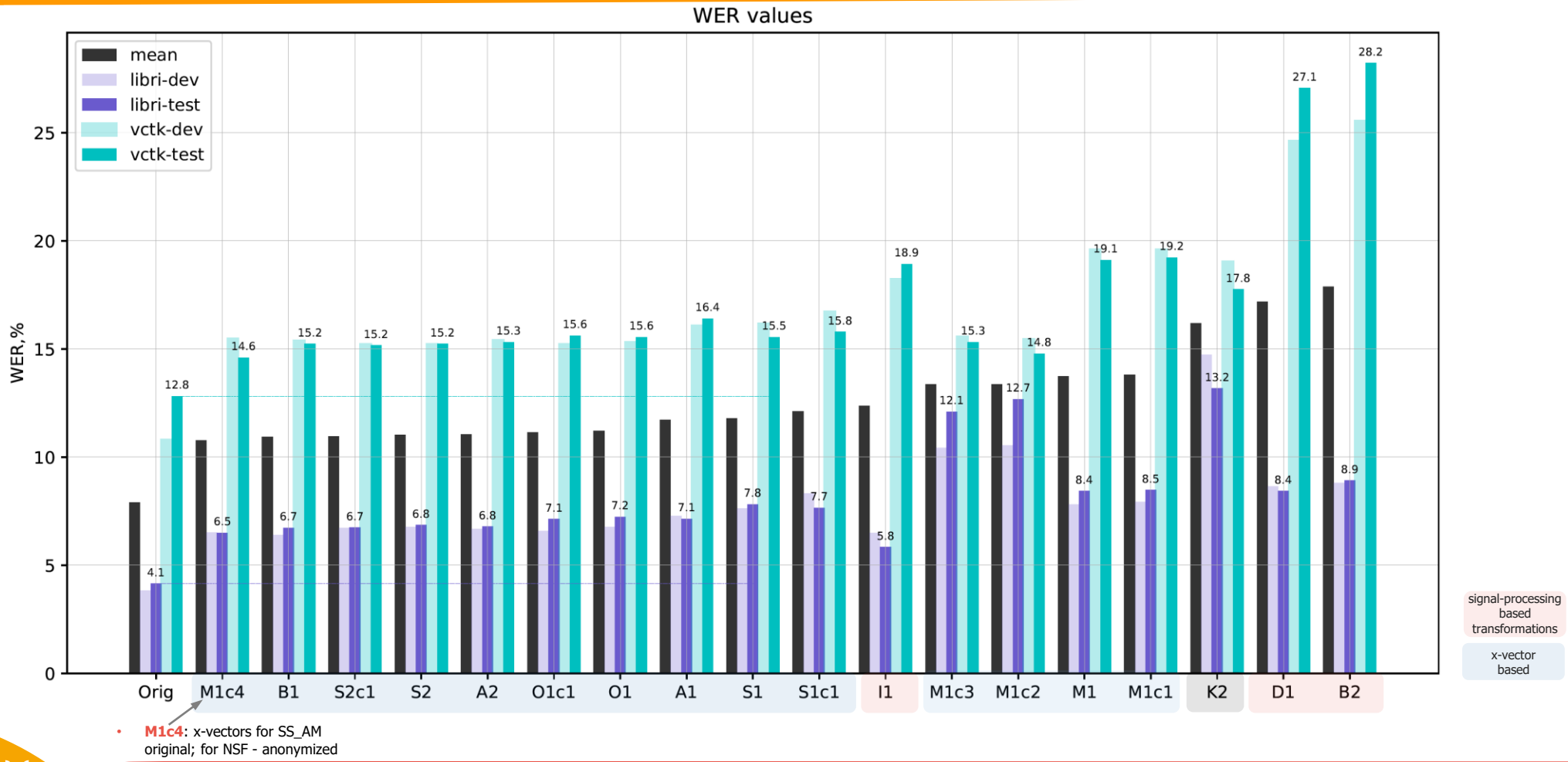
Objective evaluation results: WER for primary systems



- Anonymization incurs a large WER increase for all the systems and varies a lot depending on the method:
  - **VCTK**: **42-222%** relative
  - **LibriSpeech**: **19-120%** relative
- More WER degradation is observed on the domain-mismatch* corpus (**VCTK**)
- **Best** result for **VCTK**: **I1**
- **Best** (similar) results for **LibriSpeech**: systems using x-vector based anonymization: **B1**, **S2**, **A2**, **O1**.
- **M1** is similar to B1, but uses a different ASR model to extract BN features, that increases WER.
- In average, x-vector based anonymization methods, provides **smaller WER** than signal-processing based methods using mcAdams coefficient.

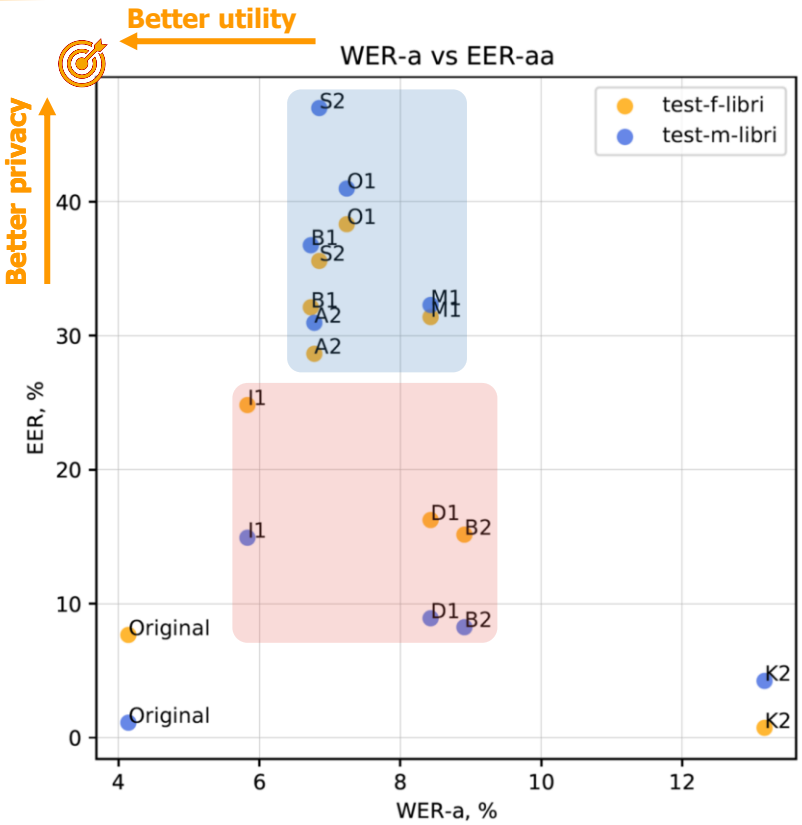
* wrt data for training ASR_eval and anonymization systems (except for D1, B2 where no training data are used)

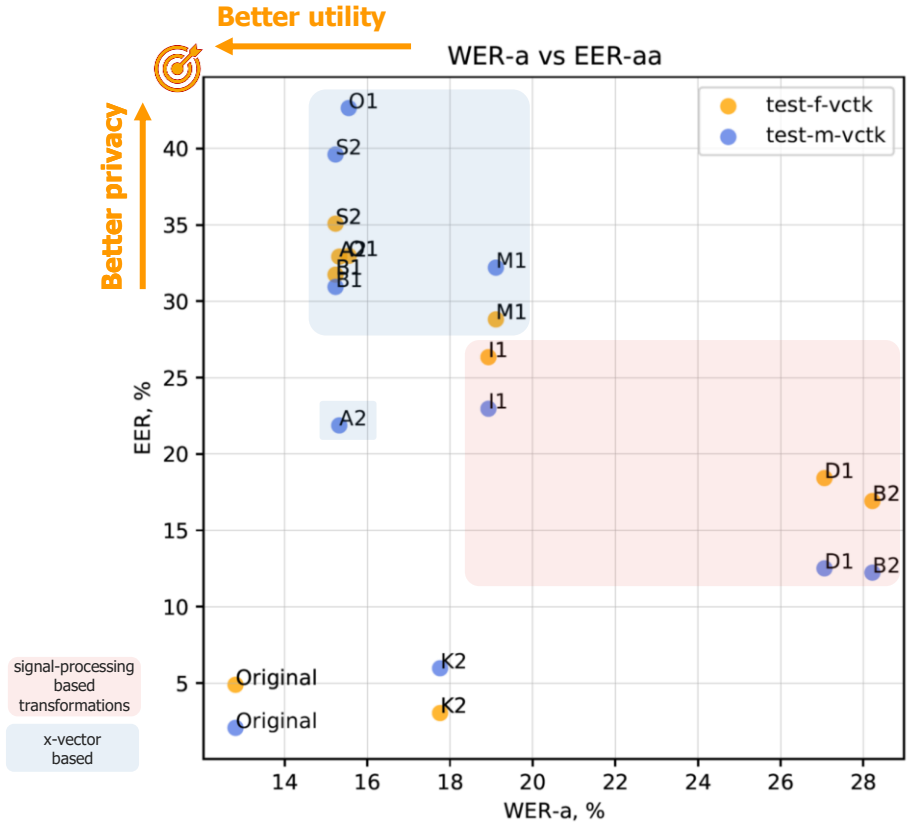
Objective evaluation results: WER for all systems



- **M1c4**: x-vectors for SS_AM original; for NSF - anonymized

Objective evaluation results: WER vs EER

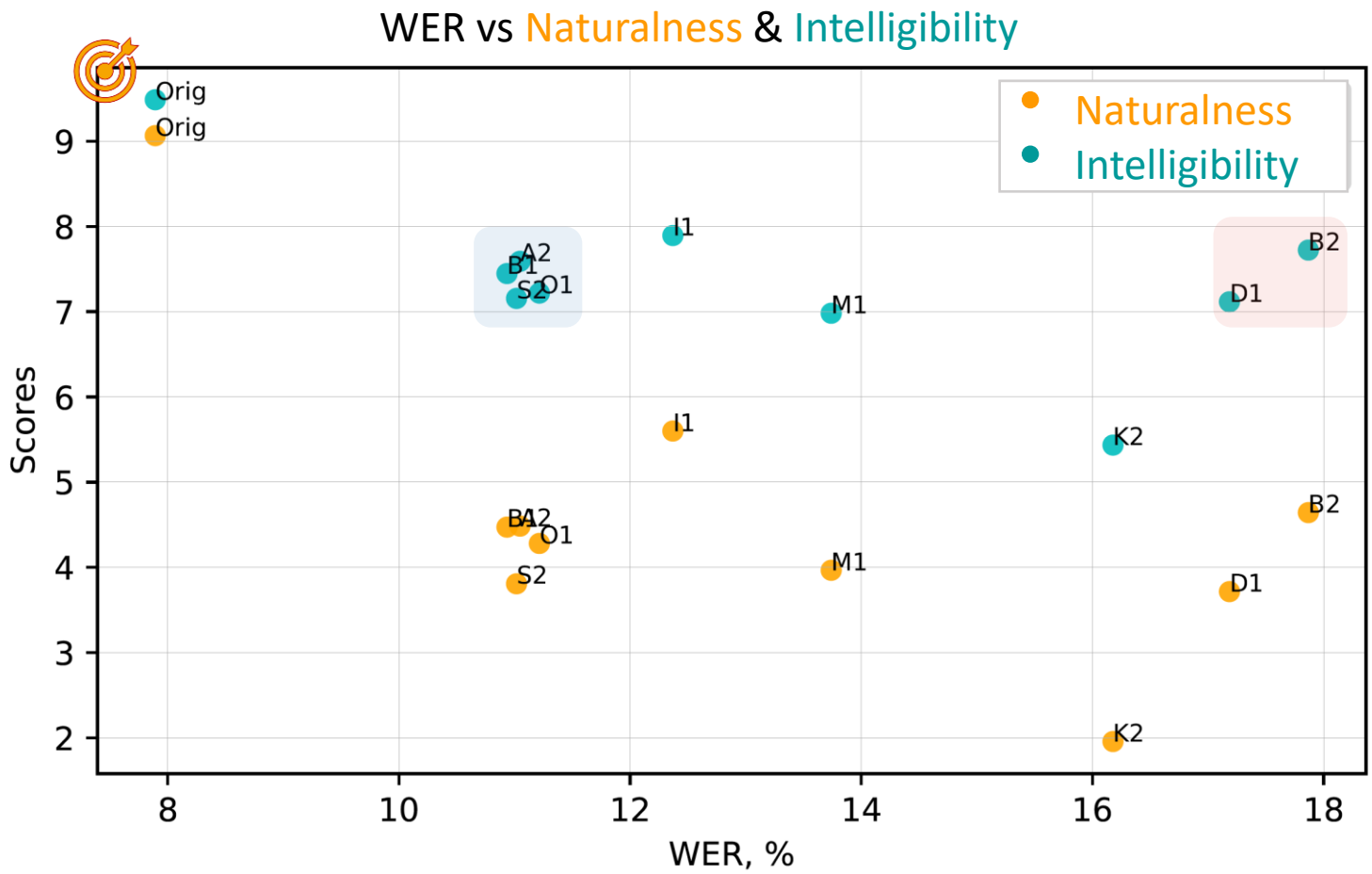




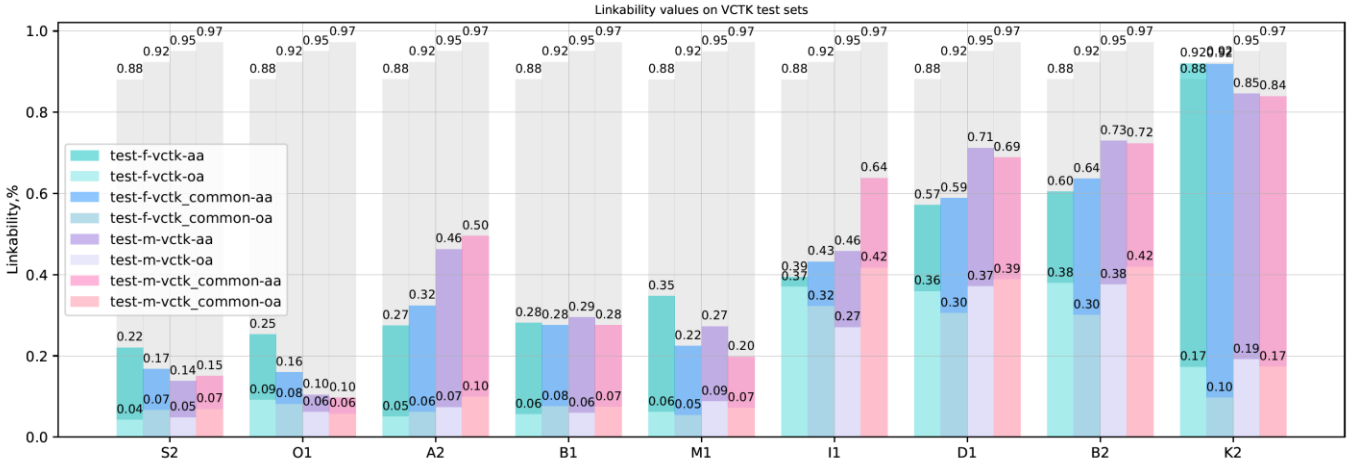
- No system which is the best for both metrics
- Best anonymization: **S2**, **O1**
- Lowest WER: **I1** (only for LibriSpeech, not stable), **B1**, **S2**, **A2**, **O1**

- **I1** provides best results among the systems using signal-processing based methods

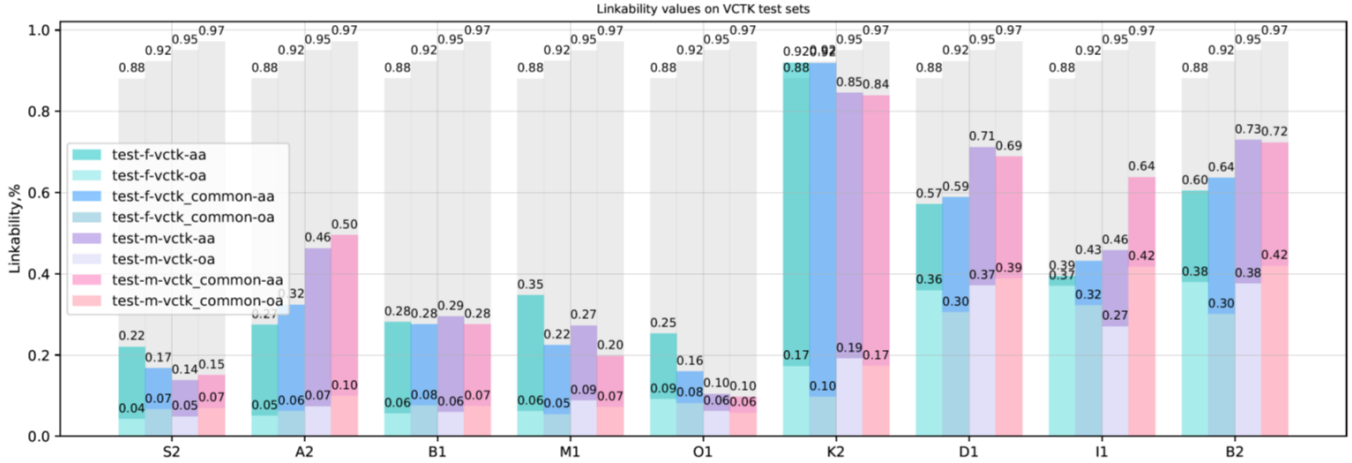
WER vs Naturalness & Intelligibility: mean scores



Linkability: VCTK-test

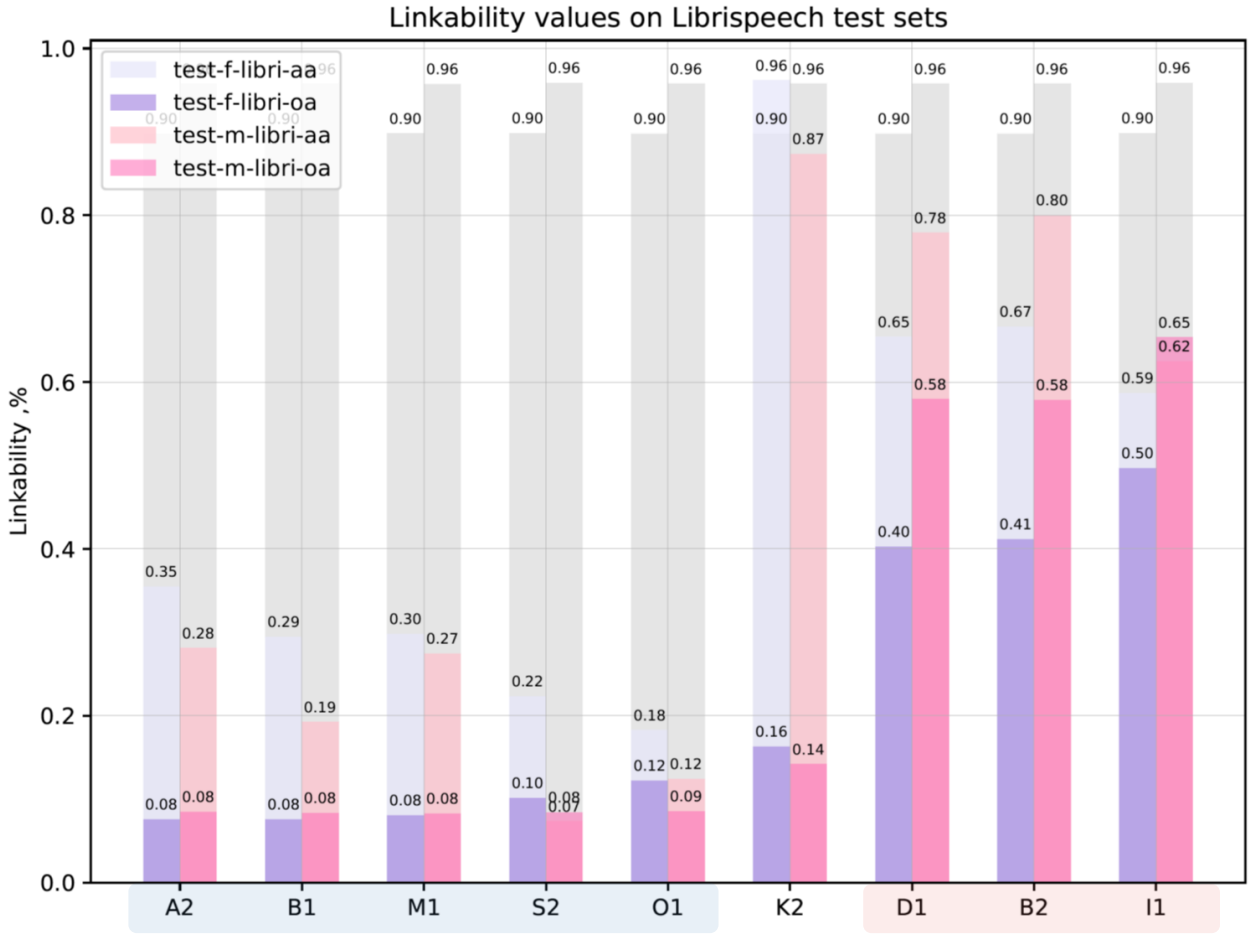


Sorted by aa test-f-vctk



Sorted by oa test-f-vctk

Linkability: Librispeech-test

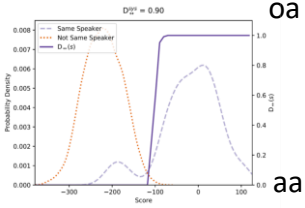


Sorted by oa test

Linkability: Librispeech-test

female:

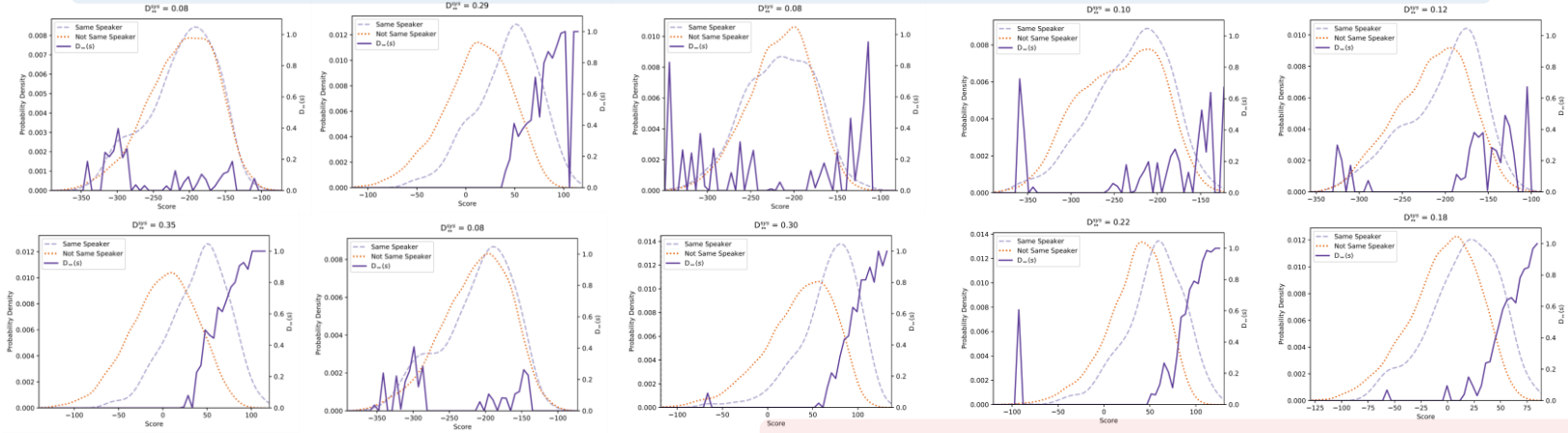
orig:



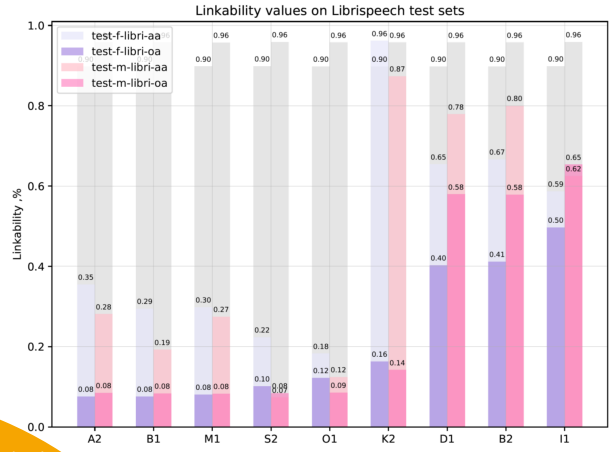
oa

aa

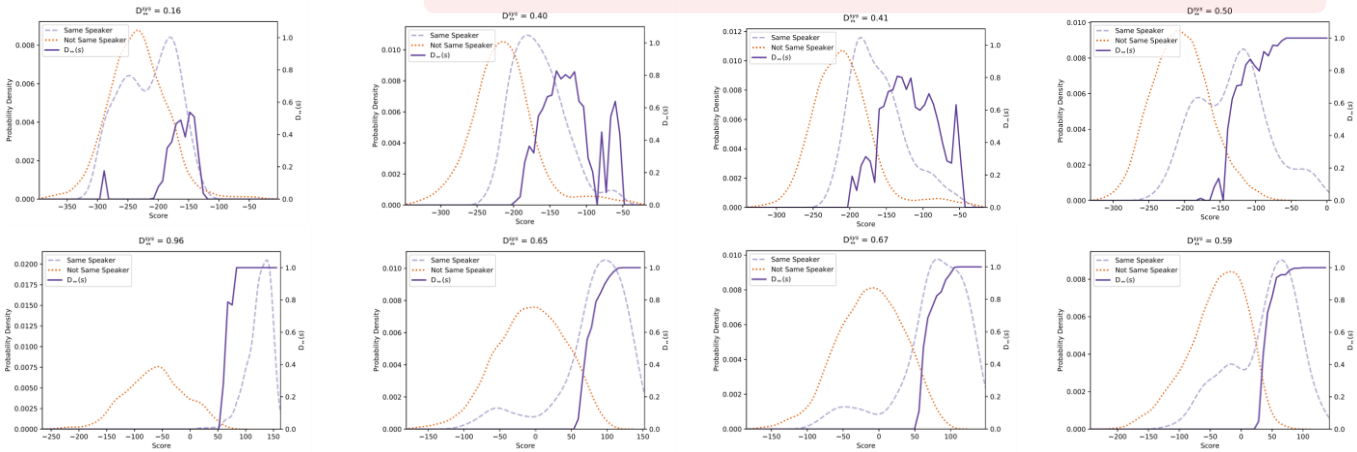
A2 B1 M1 S2 O1



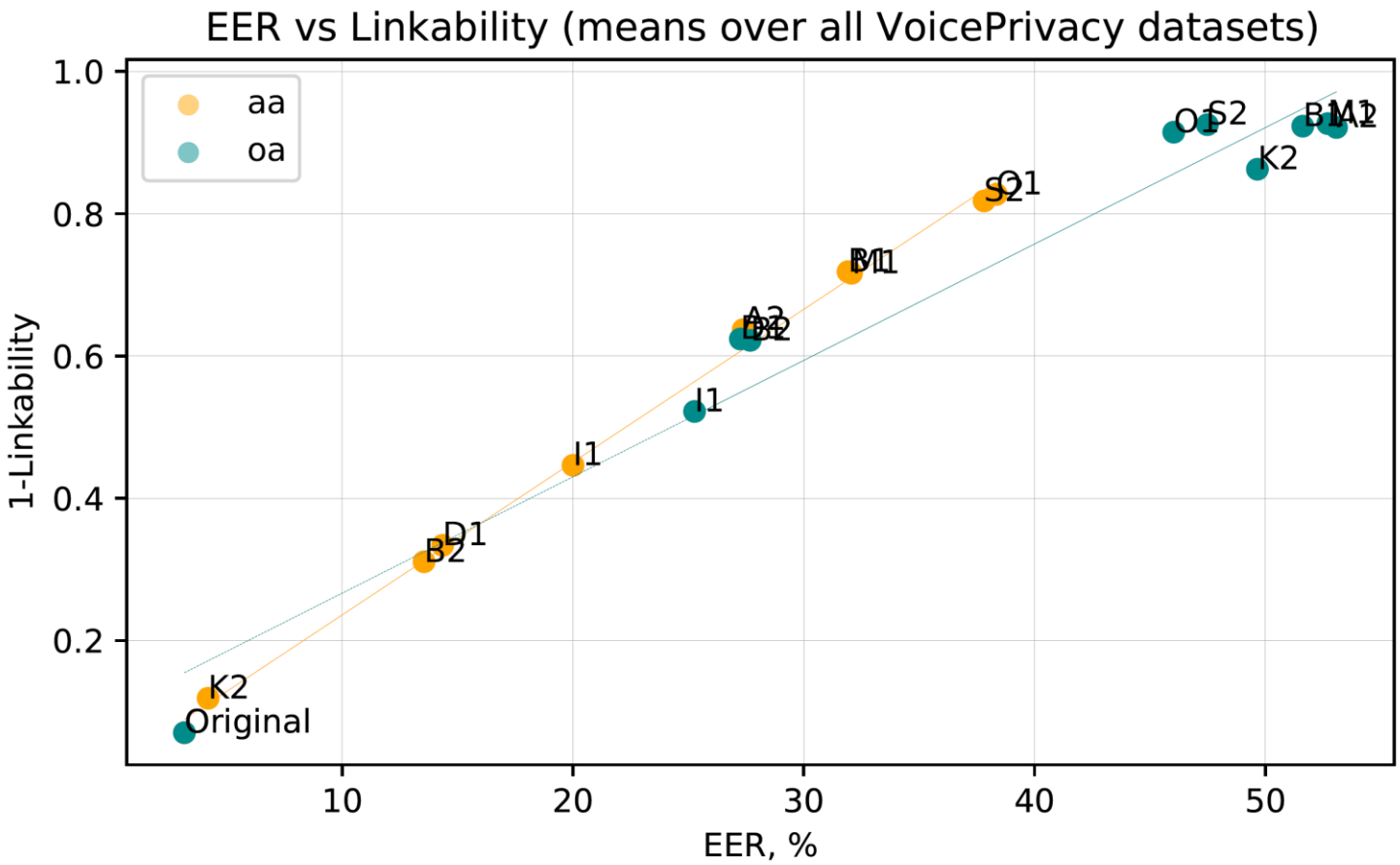
Sorted by oa test-f



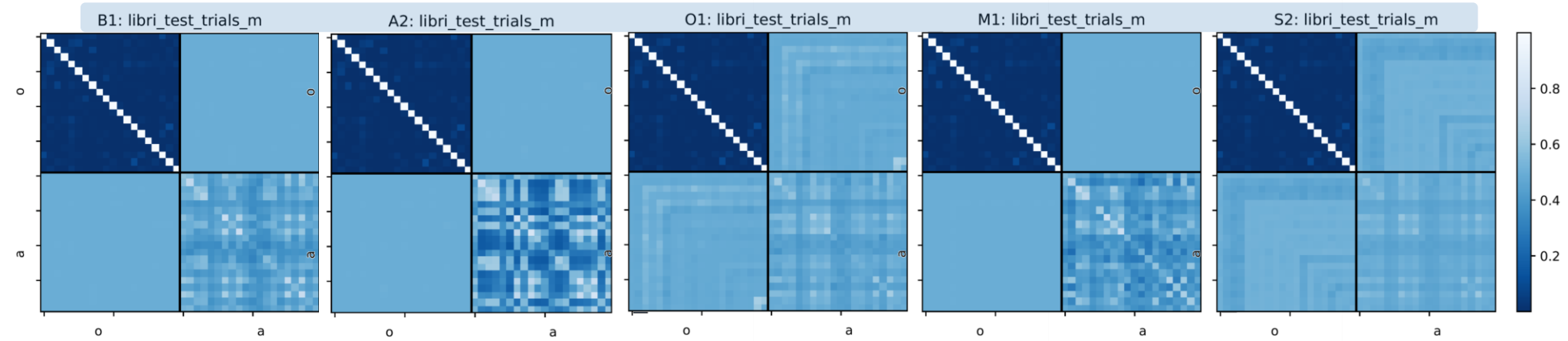
K2 D1 B2 I1

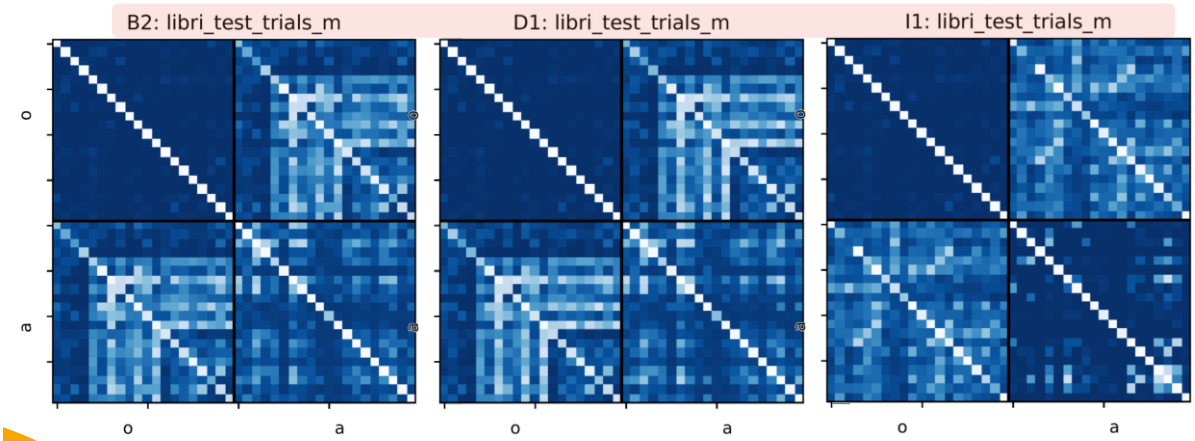


Objective evaluation results: mean EER vs Linkability



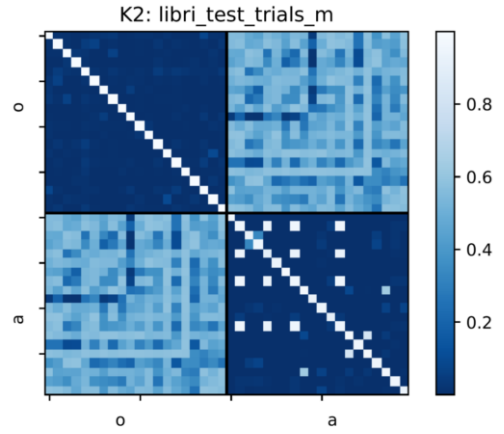
Similarity matrices: LibriSpeech-test-male



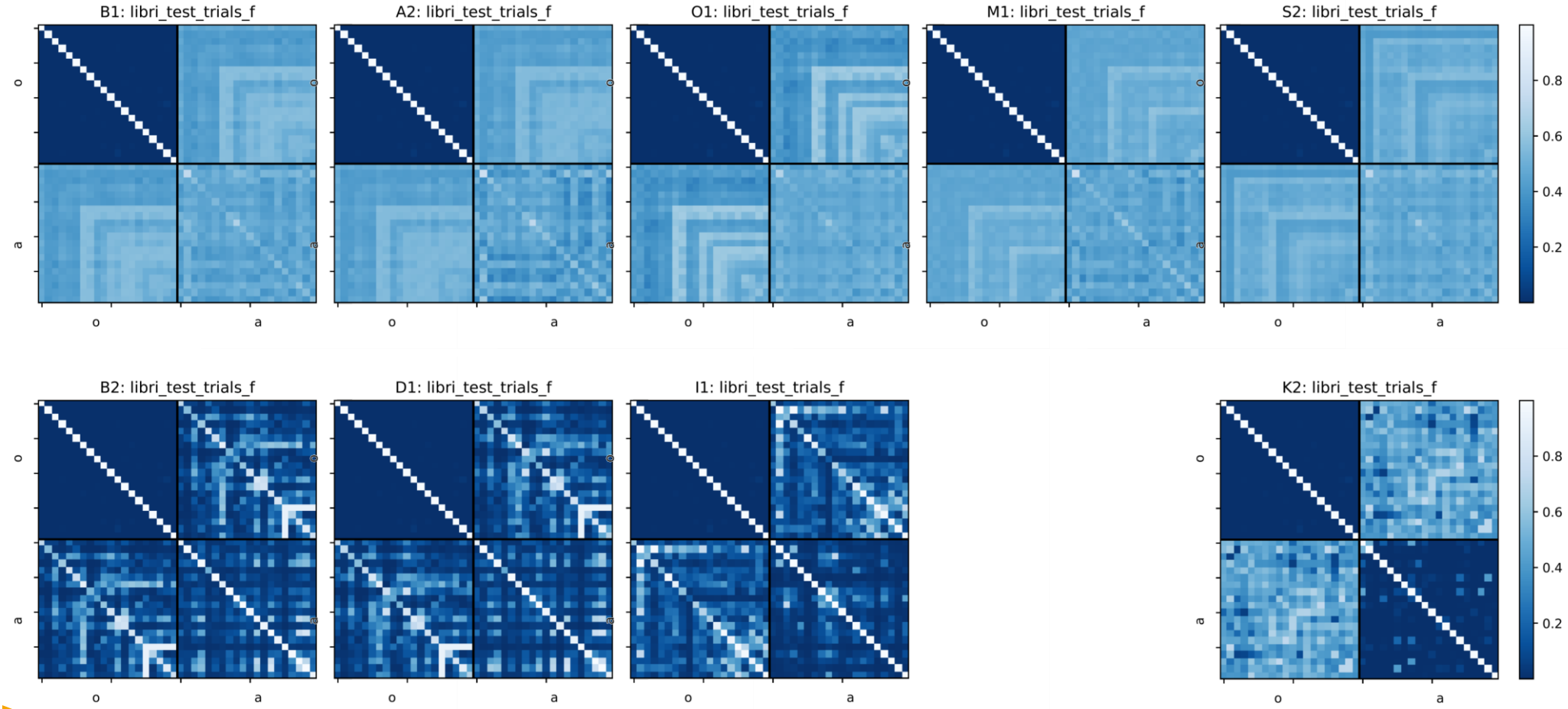


signal-processing based

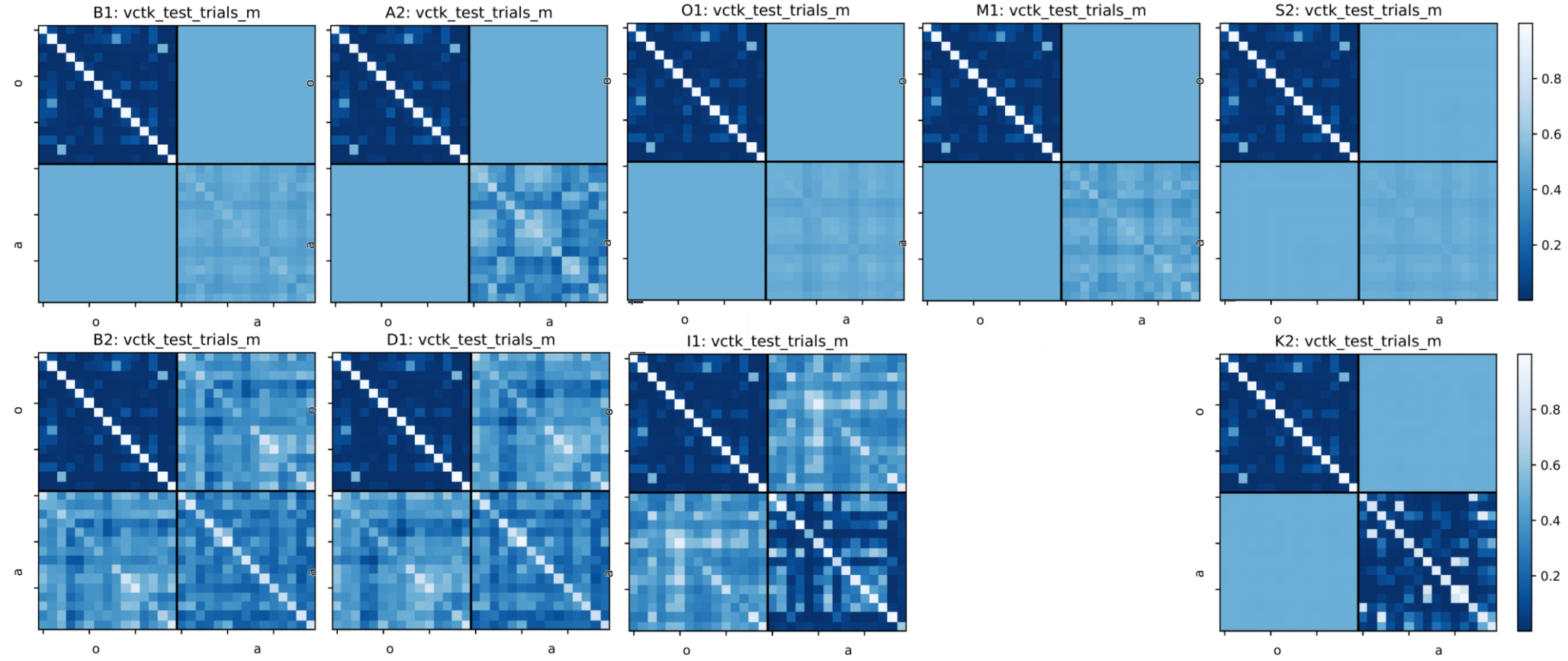
x-vector based related to B1



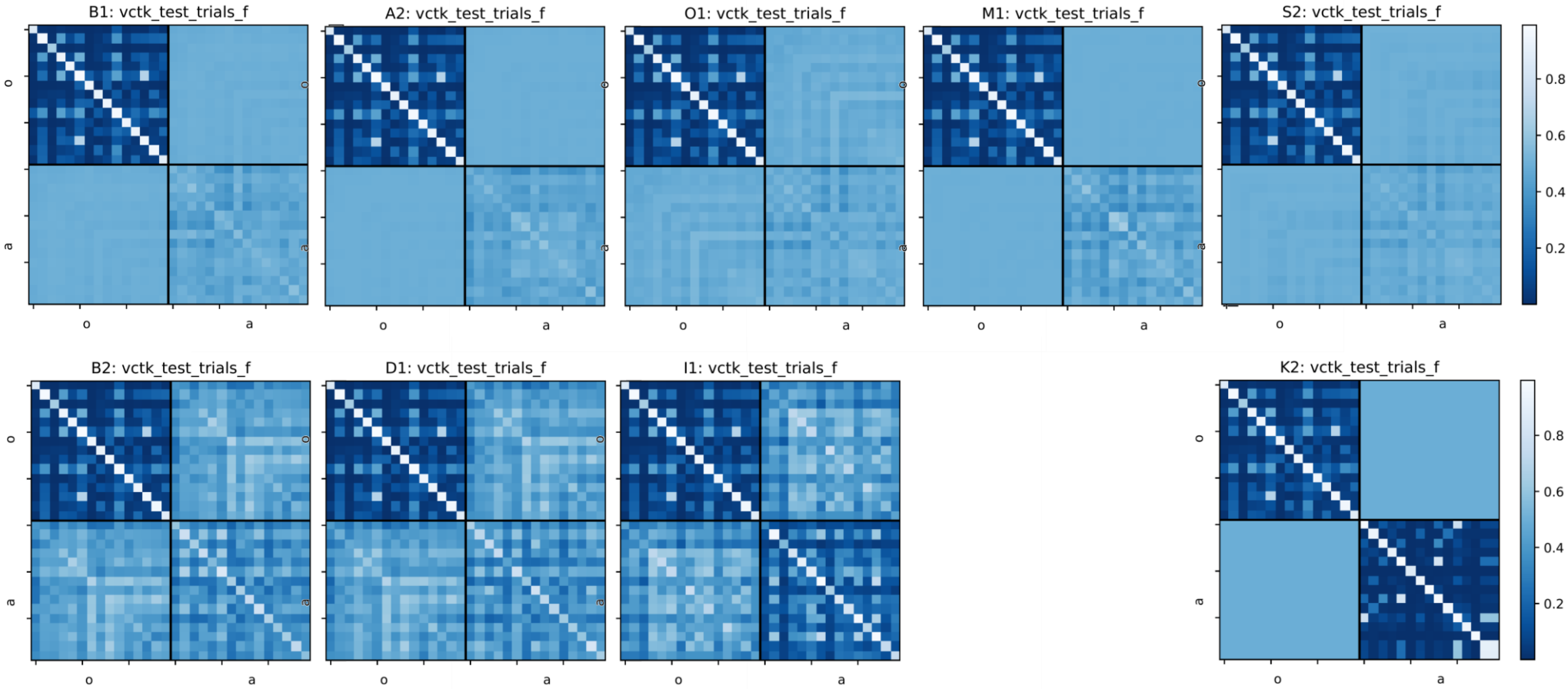
Similarity matrices: LibriSpeech-test-female



Similarity matrices: VCTK-test-male (different)



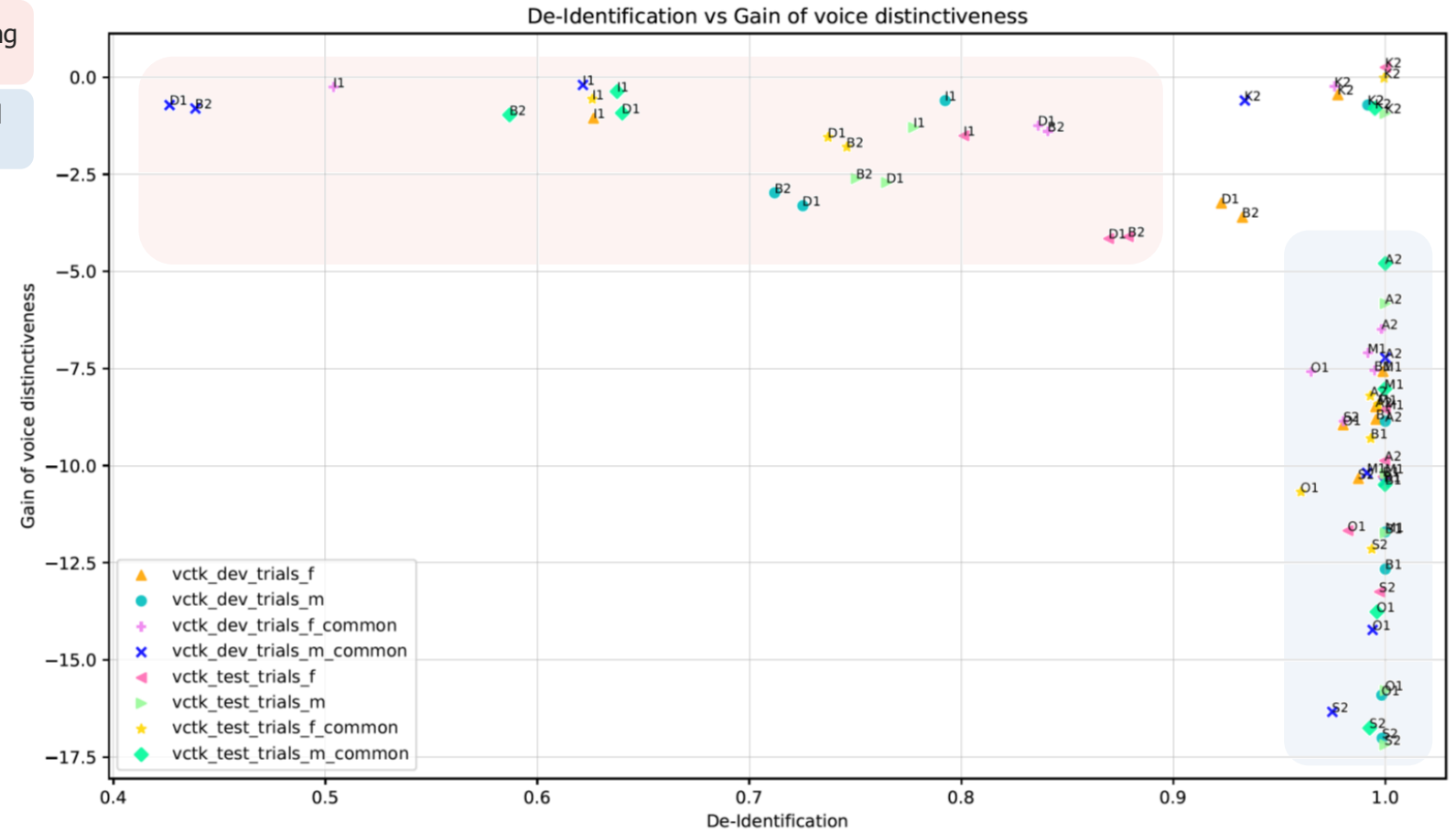
Similarity matrices: VCTK-test-female (different)



De-Identification & Gain of voice distinctiveness: VCTK

signal-processing based

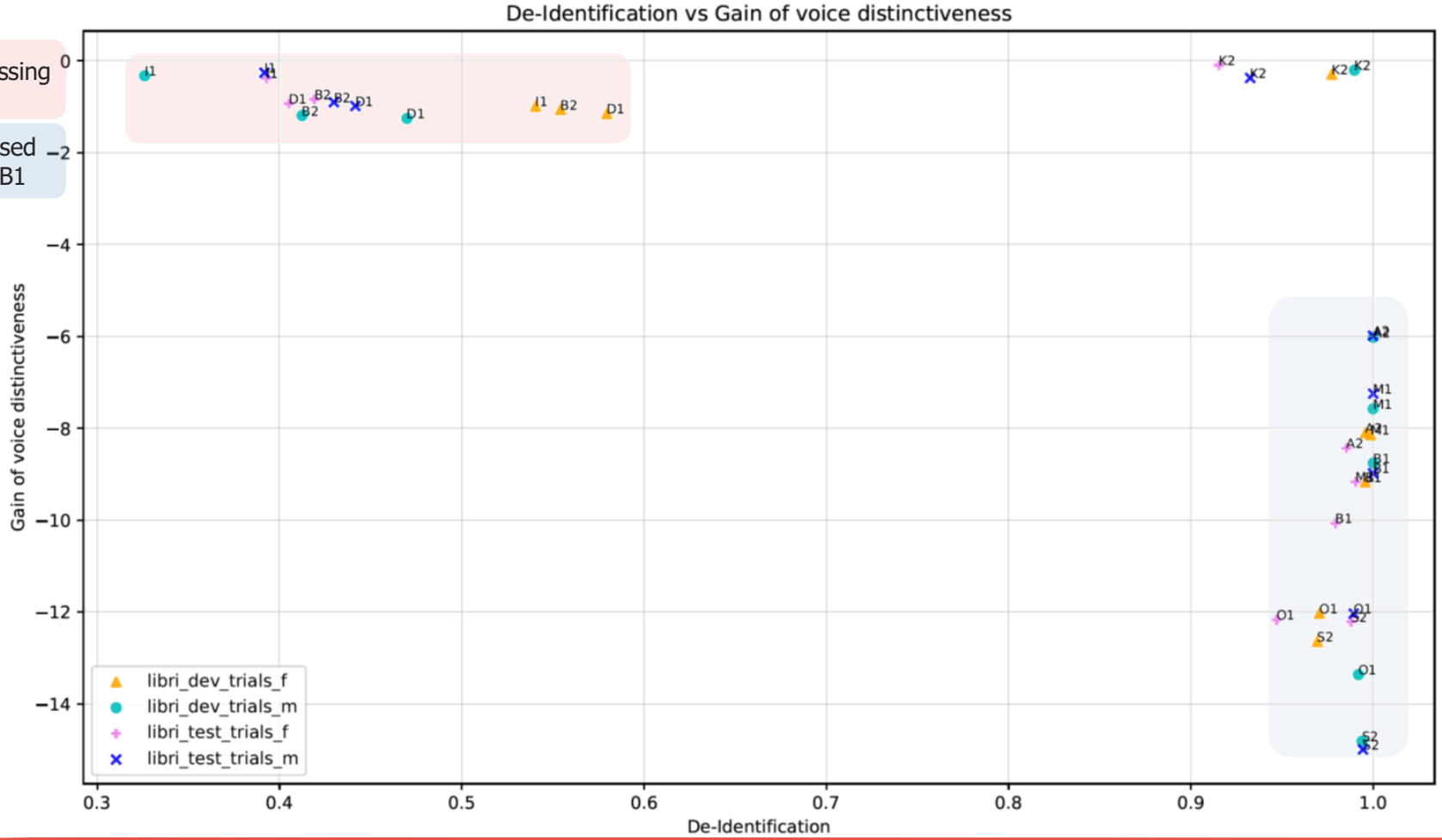
x-vector based related to B1



De-Identification & Gain of voice distinctiveness: LibriSpeech

signal-processing based

x-vector based related to B1



The VoicePrivacy 2020 Challenge

organisers@lists.voiceprivacychallenge.org

Thank you!